Degenerate U- and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency

D i s s e r t a t i o n

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der Fakultät für Mathematik und Informatik
der Friedrich-Schiller-Universität Jena

von Dipl.-Math. oec. Anne Leucht
geboren am 07.11.1983 **in** Oelsnitz (Vogtland)

Gutachter

1. Prof. Dr. Michael H. Neumann (Jena)
2. Prof. Dr. Herold Dehling (Bochum)
3. Prof. Paul Doukhan, PhD (Cergy Pontoise)

Tag der öffentlichen Verteidigung: 2011-03-04

Zusammenfassung

Ausgehend von einer stationären Folge schwach abhängiger Zufallsvariablen werden in der vorliegenden Arbeit degenerierte U- und V-Statistiken vom Grad zwei betrachtet, deren Anwendung vor allem in der mathematischen Testtheorie liegt. Eine Vielzahl von Teststatistiken, wie bespielsweise die χ^2-Statistik, die Cramér-von-Mises-Statistik sowie die Anderson-Darling-Statistik, können als degenerierte U- oder V-Statistiken formuliert beziehungsweise durch jene approximiert werden. Der erste Teil dieser Arbeit befasst sich mit der Asymptotik der vorab genannten Statistiken. Sind die zugrundeliegenden Beobachtungen unabhängig und identisch verteilt, können die Grenzverteilungen der Statistiken mittels einer Spektralzerlegung der zugehörigen Kernfunktion hergeleitet werden, sofern der Kern quadratisch integrierbar ist. Diese Methode wurde von Eagleson [56] und Carlstein [22] benutzt, um entprechende Konvergenzresultate für mischende Prozesse zu beweisen. Im gleichen Vorgehen untersuchten Dewan und Prakasa Rao [45] sowie Huang und Zhang [74] die Asymptotik bei assoziierten Variablen. Bei der Analyse all dieser Arbeiten in Kapitel 2 stellt sich heraus, dass die Übertragung der Beweisidee vom unabhängigen auf den abhängigen Fall restriktive Voraussetzungen an die aus der Spektralzerlegung resultierenden Eigenfunktionen sowie Eigenwerte erfordert. Diese Größen sind in einer Vielzahl von Anwendungen jedoch schwer oder überhaupt nicht bestimmbar und somit die an sie gestellten Bedingungen meist nicht nachweisbar.

Als erstes zentrales Ergebnis werden in Kapitel 3 dieser Arbeit die Grenzverteilungen von degenerierten U- und V-Statistiken unter leicht nachprüfbaren Voraussetzungen bestimmt. Dabei wird anstelle der Spektralzerlegung eine Waveletreihenentwicklung des Kernes vorgenommen. Bei der Verwendung dieses Ansatzes benötigt man zur Herleitung der Asymptotik lediglich Momentenbedingungen und gewisse Stetigkeitseigenschaften der Kernfunktion.

Sowohl im Falle von unabhängigen als auch bei abhängigen Beobachtungen hängen die Grenzverteilungen degenerierter U- und V-Statistiken in komplizierter Weise von gewissen Parametern ab, die wiederum auf diffizile Art von der zugrundeliegenden Situation determiniert werden. Insofern gestaltet sich die Bestimmung von (asymptotischen) kritischen Werten bei Teststatistiken vom U- beziehungsweise V-Typ problematisch. Mithilfe von Bootstrap-Verfahren können die Schwierigkeiten überwunden werden. Die Theorie zu diesen Methoden ist für degenerierte, auf unabhängigen Beobachtungen basierende U-Statistiken gut entwickelt, siehe Arcones und Giné [5], Dehling und Mikosch [40] oder Leucht und Neumann [86]. Hingegen blieb die Literaturrecherche hinsichtlich konsistenter Bootstrap-Verfahren für degenerierte U-Statistiken bei abhängigen Daten ohne Ergeb-

nis. Mit dem Ziel der Vervollkommenung der Theorie wird in Kapitel 4 der Arbeit als zweites Hauptergebnis die Konsistenz modellbasierter Bootstrap-Methoden für U- und V-Statistiken unter schwacher Abhängigkeit nachgewiesen. Dies ermöglicht es, das Anwendungspotential der Statistiken vom unabhängigen auf den abhängigen Fall zu übertragen.

In der Literatur werden zahlreiche Varianten zur Definition von *schwacher Abhängigkeit* angegeben. Am weitesten verbreitet sind Ansätze, die einen sinkenden Grad der Abhängigkeit zwischen Werten einer Zeitreihe zu vorangegangenen und zukünftigen Zeitpunkten mittels des Abfalls sogenannter Mischungskoeffizienten mit wachsender Zeitlücke zwischen „Vergangenheit" und „Zukunft" charakterisieren. Für eine Vielzahl von Prozessen können diese Eigenschaften nachgewiesen werden. Darüber hinaus lassen sich verschiedenste Instrumente der Wahrscheinlichkeitstheorie, beispielsweise zentrale Grenzwertsätze sowie Wahrscheinlichkeits- und Momentenungleichungen, vom unabhängigen Fall auf mischende Prozesse übertragen. Allerdings können bei modellbasierten Bootstrap-Verfahren Prozesse entstehen, die nicht mischend sind, obwohl die entsprechenden Originalprozesse gewissen Mischungsbedingungen genügen. Eine detaillierte Diskussion dieser Problematik erfolgt im Grundlagenkapitel 2. Es stellt sich heraus, dass für die Zwecke dieser Arbeit das alternative Konzept der *τ-dependence* besonders geeignet ist, welches von Dedecker und Prieur (2005) eingeführt wurde. Es erlaubt ein L_1-Coupling zufälliger Vektoren in folgendem Sinne: Sei X eine Zufallsvariable zu einem zukünftigen Zeitpunkt. Dann kann eine Variable $\widetilde{X}$ konstruiert werden, welche unabhängig von der „Vergangenheit" ist und die gleiche Verteilung wie X besitzt, sodass der L_1-Abstand zwischen X und $\widetilde{X}$ mit wachsender Zeitlücke zwischen „Vergangenheit" und „Zukunft" abfällt. Diese Eigenschaft wird in Abschnitt 2.1.2 mathematisch formalisert und liefert den Schlüssel zu den Beweisen in Kapitel 3 und Kapitel 4.

Um das Anwendungspotential der Theorie zu U- und V-Statistiken bei abhängigen Daten zu illustrieren, werden in Kapitel 5 drei bootstrapbasierte Hypothesentests konstruiert. Zunächst wird ein Test auf Symmetrie der Randverteilung einer Zeitreihe hergeleitet. Es folgt ein konsistentes Verfahren zur Beantwortung der Frage, ob die Randverteilung vorliegender Beobachtungen einer Zeitreihe zu einer parametrischen Verteilungsfamilie gehört. Neben Verteilungsannahmen, die mithilfe dieser beiden Tests überprüft werden können, trifft man in der Zeitreihenanalyse häufig Modellannahmen. Das heißt, es wird unterstellt, dass der Wert einer Responsevariable als Komposition einer Funktion von Informationsvariablen sowie eines zufälligen Störterms darstellbar ist. Das dritte Anwendungsbeispiel bilden auf parametrischen Bootstrap-Verfahren beruhende Tests für die Hypothese, dass die bedingte Erwartung der Responsevariablen gegeben der Informationsvariablen einer parametrischen Klasse von Funktionen angehört.

Im abschließenden Kapitel 6 wird ein Ausblick auf mögliche Verallgemeinerungen der Theorie gegeben. Die sich anknüpfenden Problemstellungen sollten in zukünftigen Forschungsprojekten verstärkt fokussiert werden.

Preface

The present thesis was written while I have been working as a research assistant at the Institute for Stochastics of the Friedrich Schiller University Jena with the aim of achieving the academic degree *Doktor der Naturwissenschaften*.

I am indebted to all those who encouraged and supported me along the way of writing this thesis. I am most grateful to my advisor Prof. Dr. Michael H. Neumann for his excellent supervision. I like to thank the PhD students, working at the institute at the same time, for the nice inspirative atmosphere. Especially, I am indebted to Barbara Wieczorek for fruitful discussions. Special thanks goes to my family who strongly encouraged me during the past three years.

Jena, 2010-12-14 Anne Leucht

Contents

Notations

Symbols

$(X_n)_{n\in\mathbb{N}}$	stationary sequence of random variables over some probability space $(\Omega, \mathcal{A}, P)$ with common distribution P_X
$\mathbb{N}$	positive integers
$\mathbb{R}_+$	positive real numbers
$\mathbb{R}^d$	space of real d-dimensional vectors, $d \in \mathbb{N}$
$\mathbb{Z}^d$	space of integer-valued d-dimensional vectors, $d \in \mathbb{N}$
$\lVert\cdot\rVert_p$	$\lVert x\rVert_p = (\sum_{i=1}^d \lvert x_i\rvert^p)^{1/p},\ x \in \mathbb{R}^d,\ p \in \mathbb{N}$
$\mathrm{Lip}(\cdot)$	Lipschitz modulus of continuity, i.e. $\mathrm{Lip}(h) = \sup_{x\neq y} \lvert h(x) - h(y)\rvert / d(x,y)$, where d is a metric and $d(x,y) = \lVert x-y\rVert_1$ unless otherwise stated
$\mathbb{1}.$	indicator function: $\mathbb{1}_A(x) = 1$ for $x \in A$ and $\mathbb{1}_A(x) = 0$ for $x \notin A$
$N(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
C	generic positive finite constant that may change its value even within a single calculation
$L_2(\mathbb{R}^d)$	$L_2(\mathbb{R}^d) := \{f : \mathbb{R}^d \to \mathbb{R} \text{ measurable } \mid \int_{\mathbb{R}^d} \lvert f(x)\rvert^2 dx < \infty\}$
$C_0(\mathbb{R}^d)$	$C_0(\mathbb{R}^d) := \{f : \mathbb{R}^d \to \mathbb{R} \text{ continuous } \mid \lim_{\lVert x\rVert_1\to\infty} f(x) \text{ exists and is equal to zero}\}$
0_d	d-dimensional null vector $(0,\dots,0)',\ d \in \mathbb{N}$
1_d	d-dimensional vector $(1,\dots,1)',\ d \in \mathbb{N}$
$\overline{A}$	closure of a set A
$\mathrm{supp}(\cdot)$	support of a function, i.e. $\mathrm{supp}(g) = \overline{\{x \mid g(x) \neq 0\}}$
$\mathbb{E}(\cdot)$	expectation
$\mathrm{var}(\cdot)$	variance
$\mathrm{cov}(\cdot,\cdot)$	covariance
$\mathbb{X}_n$	$\mathbb{X}_n = (X_1', \dots, X_n')'$
$\mathcal{B}^d$	Borel σ-algebra over $\mathbb{R}^d$
P^*	bootstrap distribution conditionally on $\mathbb{X}_n$
$\mathbb{E}^*(\cdot)$	bootstrap expectation conditionally on $\mathbb{X}_n$
$rk(\cdot)$	rank of a matrix
A'	transpose of a matrix A

$\Re(\cdot)$	real part of a complex number
$\Im(\cdot)$	imaginary part of a complex number
$\stackrel{d}{=}$	$Y \stackrel{d}{=} Z$ if the random variables Y and Z have the same distribution
$\sim$	$Z \sim Q$ if a random variable Z has distribution Q
$\Longrightarrow$	weak convergence
$\xrightarrow{d}$	convergence in distribution
$\xrightarrow{P}$	convergence in probability
$\lim$	limit
$\liminf$	limit inferior
$\limsup$	limit superior
$\forall$	for all
$\exists$	exists
$o(a_n)$	Landau notation: $b_n = o(a_n)$ if $\lim_{n\to\infty} \|b_n\|_1/\lvert a_n\rvert = 0$
$O(a_n)$	Landau notation: $b_n = O(a_n)$ if $\limsup_{n\to\infty} \|b_n\|_1/\lvert a_n\rvert < \infty$
$o_P(a_n)$	Landau notation in P-probability: $b_n = o_P(a_n)$ if $\|b_n\|_1/\lvert a_n\rvert \xrightarrow{P} 0$
$O_P(a_n)$	Landau notation in P-probability: $b_n = O_P(a_n)$ if $\forall\, \varepsilon > 0\ \exists\, K_\varepsilon \in \mathbb{R}$ such that $P(\|b_n\|_1 \geq K_\varepsilon \lvert a_n\rvert) \leq \varepsilon,\ \forall\, n \in \mathbb{N}$
$o_{P^*}(a_n)$	Landau notation in P^*-probability: $b_n = o_{P^*}(a_n)$ if $P^*(\|b_n\|_1/\lvert a_n\rvert > \varepsilon) \xrightarrow{P} 0,\ \forall\, \varepsilon > 0$
τ_r	dependence coefficient, $r \in \mathbb{N}$, see Definition 2.3 in Subsection 2.1.2

Abbreviations

a.e.	almost everywhere
a.s.	almost surely
cf.	from Latin: confer
e.g.	for example (from Latin: exempli gratia)
fidis	finite-dimensional distributions
i.e.	that is (from Latin: id est)
i.i.d.	independent and identically distributed
r.h.s.	right-hand sight
w.l.o.g.	without loss of generality
w.r.t.	with respect to

1 Introduction

Hypothesis testing constitutes one of the essential parts of mathematical statistics besides the theory of estimation. The decision for either accepting or rejecting the null hypothesis demands the evaluation of a test statistic based on a finite number of observations that are not necessarily independent of each other. In order to develop a significance test of - at least asymptotic - level α, knowledge of the limit distribution of the test statistic is required. Well-known examples include the Cramér-von Mises statistic, the Anderson-Darling statistic, and the χ^2-statistic. All three of them can be expressed in terms of degenerate V-statistics of degree two. The literature on degenerate U- and V-type statistics provides asymptotic theory for independent as well as dependent random variables. For independent and identically distributed[1] random variables, Gregory [69] derived the limit distributions via a spectral decomposition of the kernel. However, to use the same method of proof in the context of dependent data often requires restrictive assumptions whose validity is quite complicated or even impossible to check in many instances. This topic is discussed more detailed in Section 2.2.

The objective of the first part of the present thesis is the derivation of the asymptotic distributions of degenerate U- and V-statistics of weakly dependent data under easily verifiable assumptions. That is, based on a strictly stationary sequence of random variables $(X_n)_{n\in\mathbb{N}}$ over some probability space $(\Omega, \mathcal{A}, P)$ with values in $\mathbb{R}^d$, we consider

$$U_n = \frac{1}{n(n-1)} \sum_{\substack{j,k=1 \\ j\neq k}}^{n} h(X_j, X_k) \qquad \text{and} \qquad V_n = \frac{1}{n^2} \sum_{j,k=1}^{n} h(X_j, X_k).$$

Here, the function $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is called kernel of the statistics and assumed to be symmetric in its arguments. The property of degeneracy is characterized by $\int_{\mathbb{R}^d} h(x,y) P_X(dx) = 0$, where throughout the thesis P_X denotes the distribution of X_1. A wavelet decomposition of the kernel is invoked in order to determine the limit distributions of $n\,U_n$ and $n\,V_n$ in Chapter 3. This approach only requires some moment constraints and certain smoothness conditions concerning the kernel. We restrict ourselves to Lipschitz continuous kernels first. Subsequently, in Section 3.5, the latter assumption is weakened.

The asymptotic distributions for both independent and dependent observations depend on certain parameters which in turn depend on the underlying situation in a complicated way. Therefore, problems arise as soon as critical values for test statistics of U- and

[1] Throughout this thesis, the abbreviation *i.i.d.* is used instead of 'independent and identically distributed'.

V-type have to be determined. The bootstrap offers a convenient way to circumvent these difficulties. This kind of resampling technique was introduced by Efron [57] to estimate the sampling distributions of various statistics in the context of independent observations. In order to obtain validity of bootstrap methods when the random variables are (possibly) not independent, the resampling algorithm has to capture the underlying dependence structure. There are essentially two approaches to cope with this challenge. Given observations $X_1, \ldots, X_n$, the so-called block bootstrap methods do not resample single random variables but independent blocks of the underlying sample. The block length has to be chosen such that two antithetic conditions are satisfied. On the one hand, the length of the blocks has to increase with the sample size in order to capture the dependence structure properly. On the other hand, for imitating the distribution of the statistic, the number of blocks must increase with the number of observations. A comparison of different block bootstrap procedures is presented by Lahiri [83]. If the observations originate from a specific time series model, such as an AR(p) or an ARCH(p) process for instance, this additional information can be used to construct model-based methods. Since the latter approach exactly mimics the underlying dependence structure, it is expected to be more efficient than block bootstrap algorithms. A comparative portrayal of both variants is provided by Shao and Tu [100].

In the context of i.i.d. observations, the theory of bootstrap methods for degenerate U-statistics is well developed, cf. Arcones and Giné [5], Dehling and Mikosch [40], or Leucht and Neumann [86]. However, to the author's best knowledge, no results on bootstrapping general degenerate U- and V-type statistics of non-independent observations can be found in the literature. Therefore, the second main intention of this thesis is to provide sufficient conditions for deriving consistent model-based bootstrap methods for these statistics when the underlying random variables are weakly dependent. The validity of such procedures is verified in Chapter 4. Essential ingredients of the proofs are the maintenance of stationarity and of the dependence structure as well as convergence of the finite-dimensional distributions[2] of the underlying stochastic process. Additionally, certain moment constraints are required. Here, we even permit U- and V-statistics with kernels that may depend on some parameter which has to be estimated.

The last-mentioned extension is essential in order to apply the theory to hypothesis tests with composite null hypothesis. Three consistent tests of L_2-type are presented in Chapter 5. First, a test for symmetry around an unknown center is established exploiting the equivalence between symmetry of a distribution and a vanishing imaginary part of the corresponding characteristic function. The involved test statistic can be regarded as a generalization of the one proposed by Feuerverger and Mureika [67] for testing symmetry around the origin in the i.i.d. setting. Section 5.4 provides a goodness-of-fit test for the marginal distribution of a time series based on the characteristic function. This approach is eminently dedicated to cases where the probability densities do not have a closed form. A typical example is the normal inverse Gaussian distribution that is widely used in financial

[2] In this thesis, 'finite-dimensional distributions' is abbreviated by *fidis*.

mathematics, cf. Barndorff-Nielsen [8]. Finally, a fixed-kernel L_2-test for parametric model specification is considered in Section 5.5. More precisely, we are concerned with the question whether the conditional mean of some response variable given a finite-dimensional set of information, that may contain lagged values of the process under consideration, belongs to a specified parametric class. While Escanciano [60] proposed the application of wild bootstrap procedures in order to derive critical values of the test, a parametric bootstrap procedure is justified here.

Up to now we did not specify the concept of weak dependence. This is the topic of the subsequent Section 2.1. In the literature, mixing conditions are dominating in the context of quantifying the degree of dependence between random variables. Besides the fact that a great variety of processes satisfy these constraints, various tools of probability theory and statistics such as central limit theorems, probability and moment inequalities can be carried over from the i.i.d. setting to mixing processes. However, these methods of measuring dependencies are inappropriate in the present context since not only the asymptotic behaviour of U- and V-type statistics but also bootstrap consistency is focused in this thesis. Model-based bootstrap methods can yield samples that are no longer mixing even though the original sample satisfies some mixing condition, cf. Subsection 2.1.1 for a further discussion of this topic. It turns out that the characterization of dependence structures introduced by Dedecker and Prieur [35] is exceptionally suitable. Based on their τ-dependence coefficient, it is possible to construct an L_1-coupling in the following sense: Let $\mathcal{M}$ denote a σ-algebra generated by some sample variables of the past and let X be a random variable of a certain future time point. Then, the minimal L_1-distance between X and a random variable that has the same distribution as X and that is independent of $\mathcal{M}$ is equivalent to the τ-dependence coefficient $\tau(\mathcal{M}, X)$.

We exploit the coupling property in order to derive the asymptotic distributions for the original as well as the bootstrap statistics of degenerate U- and V-type. Basically, these proofs follow the same lines. First, the (almost) Lipschitz continuous kernels are approximated by a finite wavelet series expansion. There are two crucial points that assure asymptotic negligibility of the approximation error. On the one hand, the smoothness of the kernel function carries over to its wavelet approximation uniformly in scale. On the other hand, Lipschitz continuity of the kernel and the L_1-coupling property of the underlying τ-dependent sample perfectly fit together. A next step contains the application of a central limit theorem and the continuous mapping theorem to determine the limits of the approximating statistics of U-type. Based on these investigations, the asymptotic distribution of $n\,U_n$ and its bootstrap counterpart is then deduced via passage to the limit. It can be expressed as an infinite weighted sum of normal variables. Once the asymptotic distribution of $n\,U_n$ has been derived, the limit of the corresponding V-type statistic can be easily determined by applying a weak law of large numbers since

$$n\,V_n - (n-1)\,U_n = \frac{1}{n}\sum_{j=1}^{n} h(X_j, X_j).$$

The concluding Chapter 6 addresses potential extensions for future work such as the investigation of two-sample statistics as well as the consideration of block bootstrap methods for degenerate U- and V-statistics.

2 Preliminaries

This chapter establishes the basis in order to accomplish the two main results of the thesis, that is, the derivation of the asymptotic distribution of degenerate U- and V-type statistics in Chapter 3 and the verification of bootstrap consistency in Chapter 4.

In the first instance we give a brief summary of various methods to classify dependence structures. Afterwards, the concept applied in the sequel is introduced. Essential tools such as a central limit theorem for triangular schemes of random variables and a weak law of large numbers are provided.

The second part of this chapter is devoted to a survey of asymptotic results for degenerate U-statistics. The intention is to present feasible approaches to the limit distribution as well as to point out potential drawbacks of the methods available in the literature.

2.1 From independence to weak dependence

Presuming we are given a family of independent random variables, a great amount of powerful tools in probability theory and statistics, e.g. limit theorems, probability and moment inequalities, have been derived. In order to carry over these achievements and the related techniques of proof to more general situations, i.e. to cases where the assumption of independence of the underlying random variables is violated, the degree of dependence has to be somehow specified. The subsequent survey includes ordinary mixing conditions, covariance- and moment-based approaches.

Afterwards, the concept of Dedecker and Prieur [35] is elaborated on. Relationships to the aforementioned approaches are established. Finally, we will present a list of various processes whose dependence structure can be characterized by their τ-dependence coefficient.

2.1.1 Overview of concepts of weak dependence

Heuristically, a time series satisfies some weak dependence condition if the degree of dependence between observations decreases with an increasing time gap between the corresponding time points of observation. A great variety of concretisations of this property can be found in the literature.

For a long time the class of mixing processes has been dominating the mathematical formalization of weak dependence. According to the definition of specific mixing coefficients, the relation between "past" and "future" is characterized in different ways. For example, strongly mixing processes $(X_k)_{k\in\mathbb{Z}}$ are identified through

$$\alpha_k := \sup_{t\in\mathbb{Z}} \alpha(\sigma(X_s, s\le t), \sigma(X_s, s\ge t+k)) \underset{k\to\infty}{\longrightarrow} 0.$$

Here, $\sigma(X_s, s\le t)$ denotes the σ-algebra generated by $(X_s)_{s\le t}$ and, given a probability space $(\Omega, \mathcal{A}, P)$,

$$\alpha(\mathcal{U}, \mathcal{V}) := \sup_{U\in\mathcal{U},\, V\in\mathcal{V}} |P(U\cap V) - P(U)P(V)|$$

for sub-σ-algebras $\mathcal{U}$ and $\mathcal{V}$ of $\mathcal{A}$. A time series is called absolutely regular if the stricter constraint

$$\beta_k := \sup_{t\in\mathbb{Z}} \beta(\sigma(X_s, s\le t), \sigma(X_s, s\ge t+k)) \underset{k\to\infty}{\longrightarrow} 0$$

with

$$\beta(\mathcal{U}, \mathcal{V}) := \mathbb{E} \sup_{V\in\mathcal{V}} |P(V) - P(V|\mathcal{U})|$$

is fulfilled. The concept of ϕ-mixing is even more restrictive. Based on

$$\phi(\mathcal{U}, \mathcal{V}) := \sup_{\substack{V\in\mathcal{V}, U\in\mathcal{U}\\ P(U)\neq 0}} \left| P(V) - \frac{P(U\cap V)}{P(U)} \right|, \tag{2.1}$$

the condition

$$\phi_k := \sup_{t\in\mathbb{Z}} \phi(\sigma(X_s, s\le t), \sigma(X_s, s\ge t+k)) \underset{k\to\infty}{\longrightarrow} 0$$

has to be satisfied. Actually, every ϕ-mixing process is β-mixing which in turn implies the α-mixing property. For more details and the definition of further mixing coefficients we refer to the monograph of Doukhan [50]. In order to prove limit theorems and probability inequalities for mixing processes, the following coupling properties are extensively applied. An inequality associated with the α-mixing coefficient goes back to Bradley [19]. If Y is a real-valued random variable with finite moments of order $\gamma > 0$ and X is another random variable, then a random variable $\widetilde{Y} \overset{d}{=} Y$ can be constructed that is independent of X and such that

$$P\left(|\widetilde{Y} - Y| \ge q\right) \le 18 \left[\alpha^2(\sigma(X), \sigma(Y)) \frac{(\mathbb{E}|Y|^\gamma)^{1/\gamma}}{q}\right]^{\gamma/(2\gamma+1)},$$

provided that the underlying probability space is rich enough. Berbee's Lemma states a result concerning the β-mixing coefficient: Let X and Y be some random variables. Assuming that the underlying probability space is rich enough, there exists a random variable $\widetilde{Y} \overset{d}{=} Y$ that is independent of X and such that

$$P\left(\widetilde{Y} \neq Y\right) = \beta(\sigma(X), \sigma(Y)),$$

see Berbee [10]. After the introduction of the τ-dependence coefficient in the following subsection, a similar coupling property, that will play a fundamental role in the remaining part of the thesis, will be stated.

However, there are processes that are not α-mixing but still weakly dependent in the above sense. One of the most cited examples is a stationary AR(1) process with $X_t =$

$\theta X_{t-1} + \varepsilon_t$ $(0 < |\theta| \leq 1/2)$, where $(\varepsilon_t)_t$ is a sequence of independent innovations that have Binomial distribution with parameters $n = 1$ and $p = 1/2$. For a proof, see Andrews [3]. Nevertheless, it follows from $X_t = \theta^{t-s} X_s + \sum_{k=0}^{s} \theta^{t-k} \varepsilon_{t-k}$, $s < t$, that the influence of X_s on X_t decreases with increasing gap $t - s$.

More importantly, problems arise as soon as model-based bootstrap counterparts of mixing processes are investigated. For instance, it is natural to generate the bootstrap counterpart of an AR(p) process $(X_t)_t$, $p \in \mathbb{N}$, by first drawing the bootstrap innovations via Efron's bootstrap from the residuals of the original sample. After choosing a starting value X_0^*, one constructs the bootstrap AR(p) process iteratively. The difficulty of proving a mixing condition can be traced to the discreteness of the bootstrap innovations: In contrast to the distribution of its innovations, the stationary version of the bootstrap process may take values in continuous scale. Therefore, usual coupling methods to verify mixing properties for Markovian processes often fail, cf. Rosenblatt [99], Andrews [3], or Doukhan and Neumann [52]. Similar problems prompted Bickel and Bühlmann [11] to introduce a new notion of dependence, called ν-mixing condition. The associated mixing coefficient is based on the covariance of certain bounded measurable functions of a vector of finitely many "past" observations and of a vector of finitely many random variables of "future" time points.

Likewise Doukhan and Louhichi [51] proposed a covariance-based approach to measure the dependence structure within a sequence of random variables. Their motivation resulted from the difficulties to verify ordinary mixing properties and the fact that independence is equivalent to uncorrelatedness under association and for Gaussian sequences.

Definition 2.1 (Doukhan and Louhichi [51])**.** The sequence $(X_n)_{n\in\mathbb{N}}$ of random variables is called $(\theta, \mathcal{F}, \psi)$-*weak dependent*, if there exists a class $\mathcal{F}$ of real-valued functions, a sequence $\theta = (\theta_r)_{r\in\mathbb{N}}$ decreasing to zero at infinity, and a function ψ with arguments $(h, k, u, v) \in \mathcal{F}^2 \times \mathbb{N}^2$ such that for any u-tuple $(s_1, ..., s_u)$ and any v-tuple $(t_1, ..., t_v)$ with $s_1 \leq s_2 \leq ... \leq s_u < s_{u+r} \leq t_1 \leq ... \leq t_v$, one has

$$|\operatorname{cov}(h(X_{s_1}, \dots, X_{s_u}), k(X_{t_1}, \dots, X_{t_v}))| \leq \psi(h, k, u, v)\, \theta_r \tag{2.2}$$

for all functions $h, k \in \mathcal{F}$ that are defined respectively on $\mathbb{R}^u$ and $\mathbb{R}^v$.

Note that this concept is often simply called *weak dependence*. For sake of definiteness, we stick to the denotation ψ-weak dependence here. According to different specifications of the function ψ, there are various specializations of the definition above. For instance, if h and k are measurable and bounded functions and inequality (2.2) holds true with $\psi(h, k, u, v) = \mathrm{Lip}(k)\, v$, where Lip denotes the modulus of Lipschitz continuity, then this variant is called θ-weak dependence. It can be interpreted as a certain measure for causal dependence, see also Dedecker and Doukhan [29].

Numerous central limit theorems for sequences of dependent random variables can be found in the literature. Corresponding assertions under ψ-weak dependence have been derived by Doukhan and Louhichi [51], Corollary A, and by Coulon-Prieur and Doukhan

[26], Theorem 1. While the former result refers to sequences of random variables, the latter one is valid for triangular schemes of bounded random variables. A central limit theorem for triangular schemes of bounded random variables will be required when we derive the limit distributions of degenerate U- and V-statistics and their bootstrap counterparts. It turns out that the result of Coulon-Prieur and Doukhan [26] is not applicable here since they presume that the asymptotic variance behaves as in the case of independent random variables which does not hold true in our context. Therefore, a variant of the following proposition is used instead. For a certain class of ψ-weakly dependent random variables, Neumann and Paparoditis [94] verified the validity of the central limit theorem for triangular schemes.

Proposition 2.1 (Neumann and Paparoditis [94]). *Suppose that $(X_{n,k})_{k=1,\dots,n}$, $n \in \mathbb{N}$, is a triangular scheme of (row-wise) stationary random variables with $\mathbb{E}X_{n,k} = 0$ and $\mathbb{E}X_{n,k}^2 \le C < \infty$. Furthermore, we assume that*

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}X_{n,k}^2 \mathbb{1}_{|X_{n,k}|/\sqrt{n}>\varepsilon} \underset{\mathrm{n}\to\infty}{\longrightarrow} 0$$

holds for all $\varepsilon > 0$ and that

$$\frac{1}{n}\operatorname{var}\left(X_{n,1} + \dots + X_{n,n}\right) \underset{\mathrm{n}\to\infty}{\longrightarrow} \sigma^2 \in [0,\infty).$$

For $n \ge n_0$, there exists a monotonously nonincreasing and summable sequence $\{\theta_r\}_{r\in\mathbb{N}}$ such that for all indices $1 \le s_1 < \dots < s_u < s_{u+r} = t_1 \le t_2 \le n$, the following upper bounds for covariances hold true: for all measurable and quadratic integrable functions $f : \mathbb{R}^u \to \mathbb{R}$,

$$|\operatorname{cov}(f(X_{n,s_1},\dots,X_{n,s_u}), X_{n,t_1})| \le \sqrt{\mathbb{E}f^2(X_{n,s_1},\dots,X_{n,s_u})}\,\theta_r, \tag{2.3}$$

and for all measurable and bounded functions $f : \mathbb{R}^u \to \mathbb{R}$,

$$|\operatorname{cov}(f(X_{n,s_1},\dots,X_{n,s_u}), X_{n,t_1}X_{n,t_2})| \le \|f\|_\infty\,\theta_r. \tag{2.4}$$

Then,

$$\frac{1}{\sqrt{n}}(X_{n,1} + \dots + X_{n,n}) \xrightarrow{d} X \sim N(0,\sigma^2).$$

We conclude the selective overview on concepts of weak dependence with a moment-based approach recently introduced by Wu [110]. Exemplarily, a special case, called geometric-moment contraction (GMC) condition, is taken into further consideration here. This measure of dependence has been deeply discussed by Shao and Wu [102]. They investigated random variables $X_k = g(\dots,\varepsilon_{k-1},\varepsilon_k)$ where g is a measurable function and $(\varepsilon_k)_{k\in\mathbb{Z}}$ is a sequence of i.i.d. random variables. Let $(\widetilde{\varepsilon}_k)_k$ be an independent copy of $(\varepsilon_k)_k$ and define $\widetilde{X}_k = g(\dots,\widetilde{\varepsilon}_{-1},\widetilde{\varepsilon}_0,\varepsilon_1,\dots,\varepsilon_{k-1},\varepsilon_k)$. The sequence $(X_k)_k$ is said to be GMC(α) for some $\alpha > 0$ if there exist some $C > 0$ and some $\rho(\alpha) \in (0,1)$ such that

$$\mathbb{E}\|\widetilde{X}_n - X_n\|_2^\alpha \le C\,[\rho(\alpha)]^n, \quad \forall\, n \in \mathbb{N}.$$

A multitude of time series models can be verified to satisfy the constraint above, cf. Shao and Wu [102], Section 5. Examples include ARMA(p, q) models, determined through

$$X_t - \sum_{k=1}^{p} \theta_k \, X_{t-k} = \eta_t - \sum_{l=1}^{q} \phi_l \, \eta_{t-l}$$

with real-valued coefficients $(\theta_k)_{k=1}^p$, $(\phi_l)_{l=1}^q$ and innovations $(\eta_k)_{k\in\mathbb{Z}}$ satisfying GMC(α), if additionally the unit roots of $z^p - \sum_{k=1}^p \theta_k \, z^{p-k}$ lie inside the unit circle. Also GARCH(p, q) models meet GMC(α), $\alpha/2 \in \mathbb{N}$, if

$$X_t = \sqrt{h_t}\, \varepsilon_t \quad \text{and} \quad h_t = \theta_0 + \sum_{l=1}^{q} \phi_l \, X_{t-l}^2 + \sum_{k=1}^{p} \theta_k \, h_{t-k}$$

with $\mathbb{E}|X_1|^\alpha < \infty$ and if $(\varepsilon_k)_k$ is a sequence of i.i.d. centered innovations with $\mathrm{var}(\varepsilon_1) = 1$ and $\mathbb{E}|\varepsilon_1|^\alpha < \infty$.

2.1.2 τ-dependence

In the previous subsection we stated coupling properties under common mixing assumptions. Dedecker and Prieur [33] introduced a dependence coefficient that allows for an L_1-coupling for real-valued random variables. In a subsequent paper, Dedecker and Prieur [35] presented a modified version facilitating the same coupling property for random variables with values in a Polish space $(\mathcal{X}, d)$. For sake of completeness the results are presented in their whole generality, although we merely require the case $\mathcal{X} = \mathbb{R}^q$ later on.

Definition 2.2 (Dedecker and Prieur [35])**.** Let $(\Omega, \mathcal{A}, P)$ be a probability space, $\mathcal{M}$ a sub-σ-algebra of $\mathcal{A}$ and X a random variable with values in $(\mathcal{X}, d)$. Suppose that $\int d(0, x) P_X(dx) < \infty$. The *$\tau$-dependence coefficient* is defined as

$$\tau(\mathcal{M}, X) := \mathbb{E} \sup \left\{ \left| \int f(x)\, P_{X|\mathcal{M}}(dx) - \int f(x)\, P_X(dx) \right|, f \in \Lambda_1(\mathcal{X}, d) \right\},$$

where $\Lambda_1(\mathcal{X}, d)$ denotes the class of Lipschitz continuous functions $f : \mathcal{X} \to \mathbb{R}$ with $\mathrm{Lip}(f) = 1$ and $P_{X|\mathcal{M}}$ is the conditional distribution of X given $\mathcal{M}$.

This dependence coefficient has the following coupling property.

Lemma 2.1 (Dedecker et al. [30])**.** *Let $(\Omega, \mathcal{A}, P)$ be a probability space, $\mathcal{M}$ a σ-algebra of $\mathcal{A}$ and X a random variable with values in a Polish space $(\mathcal{X}, d)$. Assume that $\int d(x, x_0) P_X(dx) < \infty$ for some $x_0 \in \mathcal{X}$. Assume that there exists a random variable U, uniformly distributed over $[0, 1]$ and independent of the σ-algebra generated by $\mathcal{M}$ and X. Then there exists a random variable $\widetilde{X}$, measurable with respect to[1] $\mathcal{M} \vee \sigma(X) \vee \sigma(U)$, independent of $\mathcal{M}$ and distributed as X, such that*

$$\tau(\mathcal{M}, X) = \mathbb{E}[d(X, \widetilde{X})]. \tag{2.5}$$

[1]Throughout the thesis, *w.r.t.* abbreviates 'with respect to'.

The key to the proof of this assertion is a parametrized version of the Kantorovich-Rubinstein Theorem, cf. Dedecker et al. [30]. Note that the above coupling is optimal in the sense that $\widetilde{X}$ of Lemma 2.1 is constructed in such a manner that $\mathbb{E}(d(X,\widetilde{X}))$ is the minimal distance between X and any other variable $\bar{X}$ with $\bar{X} \stackrel{d}{=} X$ that is independent of $\mathcal{M}$ since

$$\begin{aligned}\tau(\mathcal{M},X) &= \mathbb{E}\sup\left\{\left|\mathbb{E}[f(X)\mid\mathcal{M}] - \mathbb{E}f(\bar{X})\right|, f\in\Lambda_1(\mathcal{X},d)\right\}\\ &\le \mathbb{E}(\mathbb{E}[d(X,\bar{X})\mid\mathcal{M}])\\ &= \mathbb{E}[d(X,\bar{X})].\end{aligned}\tag{2.6}$$

Applying the coupling result above, Dedecker and Prieur [34] obtained an upper bound of the τ-coefficient that involves the β-mixing coefficient. To this end, for all $y\in\mathcal{X}$ let $Q_{d(X,y)}(u) := \inf\{t\in\mathbb{R}\mid P(d(X,y)>t)\le u\}$, $u\in[0,1]$. Then, $\tau(\mathcal{M},X)\le 2\int_0^{\beta(\mathcal{M},\sigma(X))} Q_{d(X,y)}(u)\,du$, $\forall\, y\in\mathcal{X}$.

Based on $\tau(\mathcal{M},X)$, we define the causal concept of dependence that is fundamental for the remaining part of the thesis. Here, we restrict ourselves to random variables with values in $\mathbb{R}^d$ and use the metric $d(\cdot,\cdot)=\|\cdot-\cdot\|_1$ within the Definition 2.2.

Definition 2.3. A sequence of $\mathbb{R}^d$-valued integrable random variables $(X_n)_{n\in\mathbb{N}}$ over some probability space $(\Omega,\mathcal{A},P)$ is called *τ-dependent* if the sequence $(\tau_r)_{r\in\mathbb{N}}$ defined by

$$\begin{aligned}\tau_r := \sup\{&\tau\left(\sigma(X_{s_1},\dots,X_{s_u}),(X'_{t_1},X'_{t_2},X'_{t_3})'\right)\mid\\ &u\in\mathbb{N}, s_1\le\cdots\le s_u<s_u+r\le t_1\le t_2\le t_3\in\mathbb{N}\}\end{aligned}\tag{2.7}$$

satisfies $\tau_r \longrightarrow_{r\to\infty} 0$.

Remark 2.1. In the literature, the notion τ-(weak) dependence is frequently used in a slightly different sense: Given a sequence $(\mathcal{M}_i)_{i\in\mathbb{Z}}$ of sub-σ-algebras of $\mathcal{A}$, a sequence of random variables with values in a Polish space $(\mathcal{X},d)$ with $\int d(0,x)P_X(dx)<\infty$ is said to be τ-(weakly) dependent if $\sup_{k\ge0}\tau_k(r)\longrightarrow 0$ as $r\to\infty$, where

$$\tau_k(r) := \max_{1\le l\le k}\frac{1}{l}\sup_{i+r\le j_1<\cdots<j_l}\tau(\mathcal{M}_i,(X_{j_1},\dots,X_{j_l})).$$

Choosing $\mathcal{M}_i=\sigma(X_k,k\le i)$ and $\mathbb{R}^d$-valued random variables with finite expectation, our definition is less restrictive than the latter one. Although many classes of processes that satisfy (2.7) fulfil $\sup_{k\ge0}\tau_k(r)\longrightarrow 0$ as well, we stick to Definition 2.3 during the remaining part of this thesis since our slightly weaker condition suffices to prove the main results.

Besides the relation between the τ-dependence coefficient and the β-mixing coefficient, there are a lot more connections between τ-dependence in the sense of Definition 2.3 and the concepts discussed in the foregoing subsection. Of course, every m-dependent process of integrable random variables meets the definition above with $\tau_r=0$, $\forall\, r>m$. Recall that a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ is said to be m-dependent if the vectors $(X_i,\ldots,X_l)$ and $(X_{l+n},\ldots,X_j)$, $i\le\cdots\le l<l+n\le\cdots\le j\in\mathbb{Z}$, are stochastically independent whenever $n>m$.

Moreover, all integrable GMC(1) processes are τ-dependent. To verify this relation, we introduce $\widetilde{X}_{t_i}=g(\ldots,\widetilde{\varepsilon}_{s_1},\ldots,\widetilde{\varepsilon}_{s_u},\varepsilon_{s_u+1},\ldots,\varepsilon_{t_i})$, $i=1,2,3$, where the $\mathbb{R}^d$-valued function g, the variables $(X_{s_i})_{i=1}^u$, $(X_{t_i})_{i=1}^3$, and the sequence $(\widetilde{\varepsilon}_k)_k$ are defined as in the previous subsection. Obviously the vector $(\widetilde{X}_{t_1}',\widetilde{X}_{t_2}',\widetilde{X}_{t_3}')'$ is independent of $\sigma(X_{s_1},\ldots,X_{s_u})$ and has the same distribution as $(X_{t_1}',X_{t_2}',X_{t_3}')'$. Due to its construction, $(X_k)_k$ is a stationary process. Thus, the introduction of $\bar{X}_{t_i-s_u}=g(\ldots,\widetilde{\varepsilon}_{-1},\widetilde{\varepsilon}_0,\varepsilon_1,\ldots,\varepsilon_{t_i-s_u})$, $i=1,2,3$, yields

$$\begin{aligned}\mathbb{E}\left\|(\widetilde{X}_{t_1}',\widetilde{X}_{t_2}',\widetilde{X}_{t_3}')'-(X_{t_1}',X_{t_2}',X_{t_3}')'\right\|_1&=\sum_{i=1}^3\mathbb{E}\left\|\bar{X}_{t_i-s_u}-X_{t_i-s_u}\right\|_1\\&\le\sqrt{d}\sum_{i=1}^3\mathbb{E}\left\|\bar{X}_{t_i-s_u}-X_{t_i-s_u}\right\|_2\\&\le 3\,C\,\sqrt{d}\,\rho^r.\end{aligned}$$

In conjunction with Lemma 2.1 and relation (2.6), the latter inequality yields the desired structure of dependence.

The concept of τ-dependence can also be compared to a certain causal variant of ψ-weak dependence. Coulon-Prieur and Doukhan [26] introduced the s-weak dependence characterized by

$$|\mathrm{cov}\,(h(X_{s_1},\ldots,X_{s_u}),k(X_{t_1},X_{t_2}))|\le Lip(k)\,\theta_r$$

for $s_1\le s_2\le\cdots\le s_u<s_{u+r}\le t_1\le t_2$, where $(\theta_r)_{r\in\mathbb{N}}$ decreases to zero at infinity. The functions h and k are absolutely bounded by one. Additionally, k is assumed to be Lipschitz continuous. Under τ-dependence, this covariance inequality above holds with $\theta_r=\tau_r$ as

$$\begin{aligned}|\operatorname{cov}(h(X_{s_1},\ldots,X_{s_u}),k(X_{t_1},X_{t_2}))|&=|\mathbb{E}(h(X_{s_1},\ldots,X_{s_u})[k(X_{t_1},X_{t_2})-k(\widetilde{X}_{t_1},\widetilde{X}_{t_2})])|\\&\le Lip(k)\,\left[\mathbb{E}\|\widetilde{X}_{t_1}-X_{t_1}\|_1+\mathbb{E}\|\widetilde{X}_{t_2}-X_{t_2}\|_1\right].\end{aligned}$$

Here, $(\widetilde{X}_{t_1}',\widetilde{X}_{t_2}')'$ is chosen such that the assertion of Lemma 2.1 holds with $\mathcal{M}=\sigma(X_{s_1},\ldots,X_{s_u})$ and $\widetilde{X}=(\widetilde{X}_{t_1}',\widetilde{X}_{t_2}')'$, where this choice may possibly require an enlargement of the underlying probability space. Thus, τ-dependence implies s-weak dependence.

After we pointed out several relationships between τ-dependence in the sense of Definition 2.3 and well-known measures of dependence, a list of τ-dependent processes is provided below.

Example 2.1 (Causal Bernoulli shifts). Suppose that $(\varepsilon_k)_{k\in\mathbb{Z}}$ is a sequence of i.i.d. random variables and g a measurable function such that $X_k=g(\varepsilon_{k-i},i\ge 0)$ is a proper random variable for all $k\in\mathbb{Z}$, then $(X_k)_k$ forms a causal Bernoulli shift. The following processes have this representation and satisfy our τ-dependence condition.

1. We know from the previous subsection that causal ARMA(p,q) processes are GMC(1) if the corresponding sequence of innovations is GMC(1). This implies τ-dependence if the innovations are integrable. Thus, especially the stationary AR(1) process that Andrews [3] proved to be not mixing satisfies the τ-dependence condition with exponentially declining coefficients.

2. As it has also already been stated, GARCH(p,q) processes with finite second moments and square integrable innovations are GMC(2). By Lemma 2 of Wu and Min [111], the property GMC(2) implies GMC(α') for all $\alpha' \in (0,2)$, which in turn yields that those processes are τ-dependent with an exponential decay of the coefficients.

3. According to Dedecker and Prieur [35], causal linear processes with i.i.d. absolutely integrable innovations $(\varepsilon_k)_k$, i.e. $X_n = \sum_{k\geq 0} a_k\, \varepsilon_{n-k}, n \in \mathbb{N}$, satisfy (2.7) with $\tau_r \leq 6\,\mathbb{E}\varepsilon_0 \sum_{k\geq r} |a_k|$ if $\sum_{k\geq 0} |a_k| < \infty$.

Example 2.2 (Iterative random functions). 1. *Contractive models:* Let $X_k = G(X_{k-1},\dots,X_{k-p},\varepsilon_k)$, $k \in \mathbb{Z}, p \in \mathbb{N}$, where $(\varepsilon_k)_{k\in\mathbb{Z}}$ is a sequence of i.i.d. innovations and G is a measurable function with values in $\mathbb{R}^d$. Suppose that $\mathbb{E}\|G(y_0,\varepsilon_0)\|_2 < \infty$ for some $y_0 \in \mathbb{R}^{dp}$ and that there exist constants $a_1,\dots,a_p \geq 0$ with $\sum_{k=1}^p a_k < 1$ and

$$\mathbb{E}\|G(y,\varepsilon_0) - G(\bar{y},\varepsilon_0)\|_2 \leq \sum_{k=1}^p a_k \|y_k - \bar{y}_k\|_2, \quad \forall\, y = (y_1',\dots,y_p')',\ \bar{y} = (\bar{y}_1',\dots,\bar{y}_p')'.$$

Under these assumptions Shao and Wu [102] verified validity of GMC(1) for the stationary solution $(X_k)_k$ by showing that $X_k, k \in \mathbb{Z}$, has a Bernoulli shift representation. Additionally, these variables are integrable which implies that the sequence is τ-dependent with $\tau_r \leq C\,\rho^r$ for some $\rho \in (0,1)$ and a $C < \infty$.

2. *Noncontractive models:* Let $(X_k)_{k\in\mathbb{N}_0}$ be a stationary real-valued Markov chain such that $\mathbb{E}\|X_0\|_1 < \infty$ and $X_k = F(X_{k-1},\varepsilon_k)$, where $(\varepsilon_k)_{k\in\mathbb{N}}$ is a sequence of i.i.d. random variables. Then $\tau(\sigma(X_{s_1},\dots,X_{s_u}),(X_{t_1}',X_{t_2}',X_{t_3}')') \leq \sum_{i=1}^3 \mathbb{E}\|X_{t_i} - \widetilde{X}_{t_i}\|_1$ with $\widetilde{X}_k = F(\widetilde{X}_{k-1},\widetilde{\varepsilon}_k)$, where $\widetilde{X}_0 \stackrel{d}{=} X_0$ is independent of X_0, $\widetilde{\varepsilon}_k = \varepsilon_k$ for $k > s_u$ and $\widetilde{\varepsilon}_k = \bar{\varepsilon}_k$ for $k \leq s_u$. Here, $(\bar{\varepsilon}_k)_k$ is an i.i.d. copy of $(\varepsilon_k)_k$ that is also independent of X_0. In conjunction with the presumed stationarity of $(X_k)_{k\in\mathbb{N}_0}$, the Markov property leads to

$$\mathbb{E}\|X_{t_i} - \widetilde{X}_{t_i}\|_1 = \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbb{E}\|X^x_{t_i-s_u} - X^y_{t_i-s_u}\|_1 P_{X_0}(dx)P_{X_0}(dy),$$

where $(X_k^x)_{k\in\mathbb{N}_0}$ denotes the chain with starting value $X_0^x = x$. Especially if there exists a decreasing sequence $(\eta_n)_{n\in\mathbb{N}}$ such that $\eta_n \longrightarrow_{n\to\infty} 0$ and $\mathbb{E}\|X_r^x - X_r^y\|_1 \leq \eta_r\|x-y\|_1$, we obtain $\tau_r \leq 6\,\mathbb{E}\|X_0\|_1\,\eta_r,\ r \in \mathbb{N}$.

For instance, consider a stationary nonlinear AR(1) process with $X_k = f(X_{k-1}) + \varepsilon_k$, where f is a not necessarily contracting function with $f(0) = 0$ and $|f'(t)| \leq 1 -$

$C(1+|t|)^{-\delta}$ almost everywhere[2] for some $\delta \in [0,1)$ and a $C \in (0,1]$. If f is Lipschitz continuous with $\mathrm{Lip}(f) = 1$ and $\mathbb{E}|\varepsilon_1|^S < \infty, S > 1+\delta$, then $\tau_r = O(r^{(1+\delta-S)/\delta})$. For a proof, see Dedecker and Prieur [33] in conjunction with Dedecker and Rio [37].

Example 2.3 (Models with infinite memory)**.** Provided a sequence $(\varepsilon_k)_{k\in\mathbb{Z}}$ of i.i.d. random variables, Doukhan and Wintenberger [53] proved that the stationary solution of $X_t = F(X_{t-1}, X_{t-2}, \ldots; \varepsilon_t)$ almost surely[3], $t \in \mathbb{Z}$, with values in $\mathbb{R}^d$ and finite moments of order $q \in \mathbb{N}$ exists if

$$(\mathbb{E}\|F(y;\varepsilon_0) - F(\bar{y};\varepsilon_0)\|_1^q)^{1/q} \leq \sum_{j=1}^{\infty} a_j \|y_j - \bar{y}_j\|_1, \ \forall\, y, \bar{y} \in \mathbb{R}^{(\infty)}, \ a_j \geq 0, \ j \in \mathbb{N},$$

with $a = \sum_{j=1}^{\infty} a_j < 1$ and $\mathbb{E}\|F(0,0,\ldots;\varepsilon_0)\|_1^q < \infty$. Here, $x \in \mathbb{R}^{(\infty)}$ if and only if $x_k \in \mathbb{R}$ for all $k \in \mathbb{N}$ and $x_k = 0$ for all $k > N$ and some $N \in \mathbb{N}$. Invoking their Theorem 3.1, we obtain that $(X_k)_k$ is τ-dependent with

$$\tau_r \leq \frac{6}{1-a}\,\mathbb{E}\|F(0,0,\ldots;\varepsilon_0)\|_1 \inf_{1\leq p\leq r} \left[a^{r/p} + \frac{1}{1-a} \sum_{j=p+1}^{\infty} a_j \right] \underset{r\to\infty}{\longrightarrow} 0.$$

Many tools that are available for independent random variables can be transmitted to τ-dependent observations. Based on the coefficient $\tau(\mathcal{M}, X)$, Dedecker and Prieur [33, 35] derived Bennett-type inequalities as well as a functional law of iterated logarithm. Rosenthal inequalities are provided by Dedecker et al. [30]. Further results were deduced by Dedecker and Merlevède [31, 32] as well as by Dedecker and Prieur [36].

As a consequence of Proposition 2.1, we immediately obtain a central limit theorem for triangular schemes of bounded τ-dependent random variables.

Lemma 2.2. *Let $(X_{n,k})_{k=1}^n, n \in \mathbb{N}$, be a triangular scheme of (row-wise) stationary real-valued random variables with $\mathbb{E}X_{n,k} = 0$ and $\operatorname{ess\,sup}|X_{n,k}| \leq C < \infty$. Suppose that the coefficients $\bar{\tau}_r := \sup_{n>r} \tau_{r,n}$ are summable, where*

$$\begin{aligned}\tau_{r,n} := \sup\big\{\tau\left(\sigma(X_{n,s_1}, \ldots, X_{n,s_u}), (X_{n,t_1}, X_{n,t_2}, X_{n,t_3})'\right) \mid \\ u \in \mathbb{N}, 1 \leq s_1 \leq \cdots \leq s_u < s_u + r \leq t_1 \leq t_2 \leq t_3 \leq n\big\}\end{aligned} \tag{2.8}$$

with $r, n \in \mathbb{N}$. Moreover, assume that

$$\frac{1}{n}\operatorname{var}(X_{n,1} + \cdots + X_{n,n}) \underset{\mathrm{n}\to\infty}{\longrightarrow} \sigma^2 \in [0,\infty). \tag{2.9}$$

Then,

$$\frac{1}{\sqrt{n}}(X_{n,1} + \cdots + X_{n,n}) \xrightarrow{d} X \sim N(0,\sigma^2).$$

[2]Throughout the thesis, *a.e.* abbreviates ‘almost everywhere’.

[3]Instead of ‘almost surely’, the abbreviation *a.s.* will be used in the sequel.

Proof. In order to deduce the assertion from Proposition 2.1, we have to verify its prerequisites. According to stationarity and $\operatorname{ess\,sup}|X_{n,k}| \le C < \infty$, the triangular scheme exhibits uniformly bounded second moments and satisfies the Lindeberg condition. We now apply coupling techniques in order to prove (2.4). This may possibly require an enlargement of the underlying probability space. Let $f : \mathbb{R}^u \to \mathbb{R}$ be a bounded measurable function. Then, with $s_1 < \cdots < s_u < s_{u+r} = t_1 \le t_2,\ r \in \mathbb{N}$,

$$|\operatorname{cov}(f(X_{n,s_1}, ..., X_{n,s_u}), X_{n,t_1} X_{n,t_2})| \le \|f\|_\infty \, \mathbb{E}|X_{n,t_1} X_{n,t_2} - \widetilde{X}_{n,t_1} \widetilde{X}_{n,t_2}|$$

for any copy $(\widetilde{X}_{n,t_1}, \widetilde{X}_{n,t_2})'$ of $(X_{n,t_1}, X_{n,t_2})'$ that is independent of $(X_{n,s_1}, ..., X_{n,s_u})'$. In particular, the vector $(\widetilde{X}_{n,t_1}, \widetilde{X}_{n,t_2})'$ can be chosen such that

$$\mathbb{E}\big\|(X_{n,t_1}, X_{n,t_2})' - (\widetilde{X}_{n,t_1}, \widetilde{X}_{n,t_2})'\big\|_1 = \tau(\sigma(X_{n,s_1}, \ldots, X_{n,s_u}), (X_{n,t_1}, X_{n,t_2})') \le \bar{\tau}_r$$

by Lemma 2.1. This estimation implies

$$|\operatorname{cov}(f(X_{n,s_1}, ..., X_{n,s_u}), X_{n,t_1} X_{n,t_2})| \le \|f\|_\infty \, C \, \bar{\tau}_r.$$

Thus, all preliminaries of Proposition 2.1 are satisfied with $\theta_r = C\,\bar{\tau}_r,\ r \in \mathbb{N}$, with exception of inequality (2.3). The latter condition is invoked only twice within the proof of Neumann and Paparoditis [94]. It turns out that one can apply inequality (2.5) instead, under the additional assumption of uniform boundedness of the underlying triangular scheme of random variables. For sake of completeness, we state the modified calculation here. First, inequality (2.3) was used to obtain $|\operatorname{cov}(X_{n,1}, X_{n,j+1})| \le C\,\theta_j$. Here, this follows from the inequality $\mathbb{E}|X_{n,j+1} - \widetilde{X}_{n,j+1}| \le \bar{\tau}_j$ in conjunction with the uniform boundedness of the underlying random variables, where $\widetilde{X}_{n,j+1}$ is an suitably chosen copy of $X_{n,j+1}$ and independent of $X_{n,1}$. Adopting the notation of Neumann and Paparoditis [94], condition (2.3) was employed once again to derive an upper bound of order $O(n^{-1}\sum_{j=d}^{n}\theta_j)$ for the absolute value of $\Delta_{n,k}^{(1,1)} := \sum_{j=1}^{k-d} \mathbb{E} Y_{n,k} Y_{n,j} [h^{(2)}(S_{n,j} + \mu_{n,k,j} Y_{n,j} + T_{n,k}) - \mathbb{E} h^{(2)}(S_{n,k} + T_{n,k})]$, cf. page 32 of their paper. Now, let $\widetilde{X}_{n,k}$ be a copy of $X_{n,k}$ that is independent of $X_{n,1}, \ldots, X_{n,j}$ such that $\mathbb{E}|X_{n,k} - \widetilde{X}_{n,k}| \le \bar{\tau}_{k-j}$. Moreover, we introduce $\bar{T}_{n,k} \overset{d}{=} T_{n,k}$ that is independent of $(Y_{n,k})_{k=1}^n$ and $\widetilde{X}_{n,k}$. Since $h^{(2)}$ is a bounded function and $\mu_{n,k,j} = \mu(S_{n,j}, Y_{n,j}, T_{n,k})$, the following approximation holds:

$$\begin{aligned}
\left|\Delta_{n,k}^{(1,1)}\right| &= \frac{1}{\mathbb{E}(X_{n,1} + \cdots + X_{n,n})^2} \sum_{j=1}^{k-d} \mathbb{E}\big[X_{n,k} - \widetilde{X}_{n,k}\big] X_{n,j} \\
&\qquad \times \big[h^{(2)}(S_{n,j} + \mu(S_{n,j}, Y_{n,j}, \bar{T}_{n,k}) Y_{n,j} + \bar{T}_{n,k}) - \mathbb{E} h^{(2)}(S_{n,k} + T_{n,k})\big] \\
&\le \frac{C}{\mathbb{E}(X_{n,1} + \cdots + X_{n,n})^2} \sum_{j=1}^{k-d} \mathbb{E}\big|X_{n,k} - \widetilde{X}_{n,k}\big| \\
&\le O\big(n^{-1}\big) \sum_{j=1}^{k-d} \bar{\tau}_{k-j} \\
&\le O\big(n^{-1}\big) \sum_{j=d}^{n} \bar{\tau}_j.
\end{aligned}$$

Thus, with the above definition of the sequence $(\theta_r)_r$ we eventually obtain the assertion of the lemma. $\square$

Finally, we derive a weak law of large numbers for smooth functions of triangular arrays of τ-dependent random variables.

Lemma 2.3 (Weak law of large numbers). *Let $(X_{n,k})_{k=1}^n$, $n \in \mathbb{N}$, be a triangular scheme of (row-wise) stationary, $\mathbb{R}^d$-valued, integrable random variables such that $\lim_{K\to\infty}\sup_{n\in\mathbb{N}} P(\|X_{n,1}\|_1 > K) = 0$. Suppose that the coefficients $\bar{\tau}_r := \sup_{n>r}\tau_{r,n}$ satisfy $\bar{\tau}_r \longrightarrow_{r\to\infty} 0$, where*

$$\tau_{r,n} := \sup\big\{\tau\left(\sigma(X_{n,s_1},\ldots,X_{n,s_u}),(X'_{n,t_1},X'_{n,t_2},X'_{n,t_3})'\right) \mid u\in\mathbb{N},\\ 1\le s_1\le\cdots\le s_u < s_u+r\le t_1\le t_2\le t_3\le n\big\}.$$

Moreover, assume that the functions $g^{(n)}:\mathbb{R}^d\to\mathbb{R}^p$ with $\mathbb{E}g^{(n)}(X_{n,1})=0_p$ are uniformly Lipschitz continuous on any bounded interval. If additionally the sequence $(g^{(n)}(X_{n,1}))_{n\in\mathbb{N}}$ is uniformly integrable, then

$$\frac{1}{n}\sum_{k=1}^n g^{(n)}(X_{n,k}) \xrightarrow{P} 0_p.$$

Proof. Without loss of generality[4] let $p=1$. We prove that for arbitrary ε, $\eta>0$ there exists an n_0 such that for all $n>n_0$ the inequality $P\left(|n^{-1}\sum_{k=1}^n g^{(n)}(X_{n,k})|>\varepsilon\right)\le \eta$ holds. To this end, a truncation argument is invoked. Let w_K denote a Lipschitz continuous, nonnegative function that is bounded from above by one such that $w_K(x)=1$ for $x\in[-K,K]^d$ and $w_K(x)=0$ for $x\notin[-K-1,K+1]^d$ with $K\in\mathbb{R}_+$. For a finite constant M, that is specified later, define functions $g_{M,K}^{(n)}:\mathbb{R}^d\to\mathbb{R}$ by

$$g_{M,K}^{(n)}(x) := \begin{cases} g^{(n)}(x)\,w_K(x) & \text{for } |g^{(n)}(x)\,w_K(x)|\le M,\\ -M & \text{for } g^{(n)}(x)\,w_K(x) < -M,\\ M & \text{for } g^{(n)}(x)\,w_K(x) > M\end{cases}$$

and their centered versions $g_{M,K}^{(n,c)}$ by $g_{M,K}^{(n,c)}(x) = g_{M,K}^{(n)}(x) - \mathbb{E}g_{M,K}^{(n)}(X_{n,1})$. This allows for the following estimation:

$$P\Big(\Big|\frac{1}{n}\sum_{k=1}^n g^{(n)}(X_{n,k})\Big|>\varepsilon\Big) \le P\Big(\Big|\frac{1}{n}\sum_{k=1}^n g^{(n)}(X_{n,k}) - g_{M,K}^{(n)}(X_{n,k})\Big| > \frac{\varepsilon}{3}\Big)\\ + P\left(|\mathbb{E}g_{M,K}^{(n)}(X_{n,1})|>\frac{\varepsilon}{3}\right) + P\Big(\Big|\frac{1}{n}\sum_{k=1}^n g_{M,K}^{(n,c)}(X_{n,k})\Big|>\frac{\varepsilon}{3}\Big).$$

According to Markov's inequality, the first summand on the r.h.s. can be bounded by

$$\frac{3}{\varepsilon}\left[\sup_{n\in\mathbb{N}}\mathbb{E}|g^{(n)}(X_{n,1})|\mathbb{1}_{|g^{(n)}(X_{n,1})|>M} + M\sup_{n\in\mathbb{N}}P(\|X_{n,1}\|_1>K)\right].$$

[4]The expression 'without loss of generality' is abbreviated by *w.l.o.g.* in the sequel.

Since the functions $g^{(n)}$, $n \in \mathbb{N}$, are centered, we additionally obtain

$$\begin{aligned}
&P\left(|\mathbb{E}g^{(n)}_{M,K}(X_{n,1})| > \frac{\varepsilon}{3}\right) \\
&\le P\left(\sup_{n\in\mathbb{N}} \mathbb{E}|g^{(n)}_{M,K}(X_{n,1}) - g^{(n)}(X_{n,1})| > \frac{\varepsilon}{3}\right) \\
&\le P\left(\sup_{n\in\mathbb{N}} \mathbb{E}|g^{(n)}(X_{n,1})| \mathbb{1}_{|g^{(n)}(X_{n,1})|>M} + M \sup_{n\in\mathbb{N}} P(\|X_{n,1}\|_1 > K) > \frac{\varepsilon}{3}\right).
\end{aligned}$$

Therefore, by choosing M and $K = K(M)$ sufficiently large, one gets

$$P\Big(\Big|\frac{1}{n}\sum_{k=1}^{n} g^{(n)}(X_{n,k}) - g^{(n)}_{M,K}(X_{n,k})\Big| > \frac{\varepsilon}{3}\Big) + P\Big(\Big|\mathbb{E}g^{(n)}_{M,K}(X_{n,1})\Big| > \frac{\varepsilon}{3}\Big) \le \frac{\eta}{2}.$$

Concerning the remaining term, Chebyshev's inequality leads to

$$P\Big(\Big|\frac{1}{n}\sum_{k=1}^{n} g^{(n,c)}_{M,K}(X_{n,k})\Big| > \frac{\varepsilon}{3}\Big) \le \frac{9\,M^2}{\varepsilon^2 n} + \frac{18}{\varepsilon^2 n^2}\sum_{j<k} \mathbb{E}g^{(n,c)}_{M,K}(X_{n,j})g^{(n,c)}_{M,K}(X_{n,k}).$$

Thus it remains to derive an upper bound for $n^{-2}\sum_{j<k} |\mathbb{E}g^{(n,c)}_{M,K}(X_{n,j})g^{(n,c)}_{M,K}(X_{n,k})|$ that converges to zero as n tends to infinity. For this purpose we introduce a copy $\widetilde{X}_{n,k}$ of $X_{n,k}$ that is independent of $X_{n,j}$ and such that $\mathbb{E}\|X_{n,k} - \widetilde{X}_{n,k}\|_1 \le \tau_{k-j,n}$. Due to their construction, the functions $g^{(n,c)}_{M,K}$ are Lipschitz continuous uniformly in n and with a constant $C(M,K)$. To this end, note that $(g_n)_n$ is uniformly bounded on any compact interval. This implies

$$\begin{aligned}
\frac{1}{n^2}\sum_{j<k}\Big|\mathbb{E}g^{(n,c)}_{M,K}(X_{n,j})g^{(n,c)}_{M,K}(X_{n,k})\Big| &\le 2\,M\,\frac{1}{n^2}\sum_{j<k}\mathbb{E}\left|g^{(n,c)}_{M,K}(X_{n,k}) - g^{(n,c)}_{M,K}(\widetilde{X}_{n,k})\right| \\
&\le 2\,M\,C(M,K)\,\frac{1}{n}\sum_{r=1}^{n}\bar{\tau}_r,
\end{aligned} \tag{2.10}$$

where the remaining term converges to zero according to Cauchy's limit theorem, cf. Knopp [79]. □

2.2 Survey of the literature on U-statistics for dependent observations

Based on a strictly stationary sequence of random variables $(X_n)_{n\in\mathbb{N}}$ over a probability space $(\Omega, \mathcal{A}, P)$ with values in a measurable space $(\mathcal{X}, \mathcal{A}_{\mathcal{X}})$, degenerate U- and V-type statistics of degree two,

$$U_n = \frac{1}{n(n-1)}\sum_{\substack{j,k=1\\ j\neq k}}^{n} h(X_j, X_k) \qquad \text{and} \qquad V_n = \frac{1}{n^2}\sum_{j,k=1}^{n} h(X_j, X_k),$$

have been extensively investigated. The function $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called kernel and commonly assumed to satisfy the symmetry condition $h(x,y) = h(y,x)$, $\forall\, x, y \in \mathcal{X}$.

Moreover, the equation $\int_{\mathcal{X}} h(x,y)P_X(x) = 0$ holds for all $y \in \mathcal{X}$, where P_X denotes the distribution of X_1. The latter property is referred to as degeneracy of the kernel and the associated statistic, respectively. Extensions to kernels that map in more general spaces were for instance considered by Dehling, Denker and Philipp [39]. Here, we restrict ourselves to real-valued kernel functions.

There are essentially two approaches to derive the asymptotic distributions of the statistics $n\,U_n$ and $n\,V_n$. In the following subsection we give an overview of limit theorems that extend the result of Gregory [69] who invoked a spectral decomposition of the function h to establish the asymptotics of degenerate U-statistics of i.i.d. random variables under contiguous alternatives. Then again, one can understand U- and V-statistics as integrals w.r.t. the empirical process. Corresponding limit results on $n\,U_n$ and $n\,V_n$ based on empirical process theory are recapitulated in Subsection 2.2.2.

2.2.1 A first approach: Spectral decomposition of the kernel

A spectral decomposition of the kernel constitutes the starting point of various papers on the asymptotics of degenerate U- statistics. Below we sketch the method of Gregory [69] for deriving the limit distribution of $n\,U_n$ in order to motivate our approach that is presented in Chapter 3.

For this purpose the function h is assumed to satisfy $\iint_{\mathcal{X}\times\mathcal{X}} h^2(x,y)\,P_X(dx)\,P_X(dy) < \infty$. Now, the integral equation

$$\lambda\,q(x) = \int_{\mathcal{X}} h(x,y)q(y)P_X(dy) \tag{2.11}$$

is considered. We denote the corresponding, possibly finite sequence of orthonormal eigenfunctions by $(\Phi_k)_k$ and the associated eigenvalues by $(\lambda_k)_k$. The term orthonormality has to be understood in the sense of $\mathbb{E}[\Phi_j(X_1)\Phi_k(X_1)] = \delta_{j,k}$, where $\delta_{j,k}$ denotes the Kronecker delta. The theory on Hilbert-Schmidt operators yields

$$\lim_{K\to\infty} \iint_{\mathcal{X}\times\mathcal{X}} \left[h(x,y) - \sum_{k=1}^{K} \lambda_k\,\Phi_k(x)\,\Phi_k(y)\right]^2 P_X(dx)\,P_X(dy) = 0,$$

which will be denoted shortly by

$$h(x,y) = \sum_{k=1}^{\infty} \lambda_k\,\Phi_k(x)\,\Phi_k(y), \tag{2.12}$$

see e.g. Dunford and Schwartz [55], Section XI.6. This relation suggests to approximate $n\,U_n$ by

$$n\,U_{n,K} := \frac{n}{n-1}\sum_{k=1}^{K}\lambda_k\left[\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Phi_k(X_j)\right)^2 - \frac{1}{n}\sum_{j=1}^{n}\Phi_k^2(X_j)\right].$$

In order to obtain the limit distribution of $n\,U_{n,K}$ as n tends to infinity, the strong law of large numbers is applied to the subtrahend. Additionally, the central limit theorem yields that the expression in the round brackets tends to a standard normal random variable. To this end, note that $\mathbb{E}\Phi_k(X_1) = 0$ for all k with $\lambda_k \neq 0$. Furthermore, one can show that

$$\iint_{\mathcal{X}\times\mathcal{X}} \Big[h(x,y) - \sum_{k=1}^{K} \lambda_k \Phi_k(x)\Phi_k(y)\Big]^2 P_X(dx)P_X(dy)$$
$$= \iint_{X\times X} h^2(x,y)P_X(dx)P_X(dy) - \sum_{k=1}^{K} \lambda_k^2,$$

which in turn leads to $\sum_{k=1}^{\infty} \lambda_k^2 = \iint_{\mathcal{X}\times\mathcal{X}} h^2(x,y)P_X(dx)P_X(dy) < \infty$. Moreover, under the assumption $\mathbb{E}|h(X_1,X_1)| < \infty$, the strong law of large numbers yields

$$n\,V_n = \frac{n-1}{n}[n\,U_n] + \frac{1}{n}\sum_{j=1}^{n} h(X_j,X_j) = \frac{n-1}{n}[n\,U_n] + \mathbb{E}h(X_1,X_1) + o_P(1).$$

Eventually, these are the main steps to prove the following theorem, cf. Lee [84], Sections 3.2.2 and 4.2.

Theorem 2.1. *Suppose that $(X_n)_{n\in\mathbb{N}}$ is a sequence of i.i.d. random variables. Let U_n be a degenerate U-statistic with kernel h satisfying $\mathbb{E}h^2(X_1,X_2) < \infty$. Then, as $n \to \infty$,*

$$n\,U_n \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1),$$

where $(Z_k)_k$ is a sequence of independent standard normal variables. Under the additional condition $\mathbb{E}|h(X_1,X_1)| < \infty$, the corresponding sequence of degenerate V-type statistics satisfies

$$n\,V_n \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1) + \mathbb{E}h(X_1,X_1).$$

Note that the sum determining the limit variables converges in the L_2-sense as $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$. So far, most previous attempts to derive the limit distributions of degenerate U- and V-statistics of dependent random variables are based on the adaptation of the method of proof we sketched above.

Eagleson [56] considered stationary sequences of ϕ-mixing, real-valued random variables. Concerning the decay of the mixing coefficients, he assumed $\sum_{k=1}^{\infty} \phi_k^{1/2} < \infty$. If the kernel of a degenerate U-statistic is square integrable w.r.t. $P_X \times P_X$ and if $\sum_{k=1}^{\infty} |\lambda_k| < \infty$, then

$$n\,U_n \xrightarrow{d} U := \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1). \tag{2.13}$$

Here, $(Z_k)_{k\in\mathbb{N}}$ is a sequence of centered jointly normal random variables with $\operatorname{cov}(Z_j,Z_k) = \mathbb{E}\Phi_j(X_1)\Phi_k(X_1) + \sum_{i=2}^{\infty}[\mathbb{E}\Phi_j(X_1)\Phi_k(X_i) + \mathbb{E}\Phi_j(X_i)\Phi_k(X_1)]$.

Remark 2.2. (i) In general, it is rather difficult to establish easily verifiable conditions on the function h that ensure absolute summability of the eigenvalues. However, there exist examples of kernels that allow for an explicit calculation of the associated eigenvalues, see e.g. Darling [27], Moore and Spruill [90], or Dehling [38]. Exemplarily, we consider the generalized Cramér-von Mises statistic

$$T_n = n \int_{\mathbb{R}} (F_n(x) - F(x))^2 \, w(F(x)) \, dF(x), \tag{2.14}$$

which is applied to test whether the underlying observations have the distribution function F. Here, F_n denotes the empirical distribution function, i.e. $F_n(x) = n^{-1} \sum_{k=1}^{n} \mathbb{1}_{X_k \leq x}$, $x \in \mathbb{R}$. Obviously, T_n can be expressed in terms of a V-statistic with kernel

$$h(x, y) = \int_{-\infty}^{\infty} (\mathbb{1}_{x \leq u} - F(u)) \, (\mathbb{1}_{y \leq u} - F(u)) \; w(F(u)) \, dF(u)$$

In the case of the ordinary Cramér-von Mises statistic, the weight function satisfies $w \equiv 1$. Here, the associated eigenvalues of (2.11) are given by $\lambda_k = (k\pi)^{-2}$, $k \in \mathbb{N}$. Thus, the condition $\sum_{k=1}^{\infty} |\lambda_k| < \infty$ is satisfied. It also holds true in the trivial case of only finitely many non-zero eigenvalues. In particular, the χ^2-statistic of order k exhibits $k - 1$ non-vanishing eigenvalues, cf. de Wet [47] and references therein.

(ii) According to Dunford and Schwartz [55], Section XI.6, the eigenvalues of kernel functions h defined by $h(x, y) = \int_{\mathbb{R}} h_1(x, z) h_2(z, y) P_X(dz)$ are absolutely summable if $\mathbb{E}|h(X_1, X_1)| < \infty$ and h_1 as well as h_2 are square integrable w.r.t. $P_X \times P_X$. These types of kernels occur in the context of L_2-type statistics. The integrability condition is violated for instance if the weight function in (2.14) is heavy tailed. An example, namely $w(u) = \varphi\left(\Phi^{-1}(u)\right)^{-2}$, is provided by de Wet [47]. Here, φ denotes the density and Φ the cumulative distribution function of the standard normal distribution. In this case the eigenvalues are given by $\lambda_k = 1/k$, $k \in \mathbb{N}$, and are therefore not summable.

Carlstein [22] investigated U-statistics with square integrable kernels (w.r.t. $P_X \times P_X$) of stationary $*$- and α-mixing sequences of $\mathbb{R}^d$-valued random variables. A stationary process $(X_k)_{k \in \mathbb{Z}}$ is said to be $*$-mixing if

$$\begin{aligned} \psi_r := \sup\{ &|P(U \cap V) - P(U)P(V)| / [P(U)P(V)] \,|, P(U) \neq 0 \neq P(V) \\ &U \in \sigma(X_s, s \leq t), \, V \in \sigma(X_s, s \geq t + r)\} \end{aligned}$$

decreases to zero as r tends to infinity. Under the additional condition of absolutely summable eigenvalues and if $\sum_{r=0}^{\infty} (r + 1)\sqrt{\psi_r} < \infty$, he obtained the limit distribution of $n \, U_n$. In the case of an α-mixing sequence with $\sum_{k=1}^{\infty} \alpha_k^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$, Carlstein assumed the joint distribution of any two variables of the sequence to be absolutely continuous w.r.t. the product measure of the marginals. Concerning the kernel of U_n, he postulated that its orthogonal expansion yields only finitely many non-vanishing eigenvalues. The associated eigenfunctions were supposed to satisfy $\mathbb{E}|\Phi_k(X_1)|^{2+\delta} < \infty$. Under these constraints, he obtained $n\,U_n \xrightarrow{d} U$, where U is defined as in (2.13).

An analogous result was established by Chen and White [25]. They considered a stationary sequence $(X_k)_{k\in\mathbb{Z}}$ of integrable random variables with values in a Hilbert space $(\mathcal{X}, d)$ with norm $\|\cdot\|$ such that there exists a compact set $A \subset \mathcal{X}$ with $P(X_1 \in A) = 1$. The sequence was supposed to be near epoch dependent on a strongly mixing process $(V_l)_{l\in\mathbb{Z}}$ with coefficients $\alpha_k = O(k^{-2r/(r-2)})$ for some $r > 2$. More precisely, the existence of nonnegative constants $(d_i)_{i\in\mathbb{N}}$ and $(\mu_m)_{m\in\mathbb{N}}$ with $\mu_m = O(m^{-1-\delta})$ for some $\delta > 0$ such that $(\mathbb{E}\|X_i - \mathbb{E}[X_i \mid \sigma(V_l,\, i-m \leq l \leq i+m)]\|^2)^{1/2} \leq \mu_m\, d_i$ was presumed. Regarding the kernel of the degenerate U-statistic, they assumed $\sup_{k\in\mathbb{N}} \mathbb{E}|h(X_1, X_k)|^{2+\delta} + \iint_{\mathcal{X}\times\mathcal{X}} |h(x,y)|^{2+\delta} P_X(dx)P_X(dy) < \infty$ and absolute summability of the eigenvalues of (2.11). Again, the limit distribution of $n\,U_n$ is equivalent to the distribution of U, defined in (2.13).

Denker [43] obtained the limit distributions of degenerate U- and V-statistics for stationary sequences of real-valued functionals $X_n = f(Z_n, Z_{n+1}, \dots)$, $n \in \mathbb{N}$, of β-mixing random variables $(Z_n)_{n\in\mathbb{Z}}$ with $\beta_k = O(k^{-8})$. He assumed f and the cumulative distribution function F of X_1 to be Hölder continuous. Denoting the Hölder exponent of F by r, Denker imposed the following smoothness condition on the kernel function:

$$\sup_{\varepsilon>0} \varepsilon^{-r} \iint_{\mathbb{R}\times\mathbb{R}} \sup_{\substack{|x-u_1|+|y-u_2|\leq\varepsilon; \\ |\bar{x}-u_1|+|\bar{y}-u_2|\leq\varepsilon}} |h(x,y) - h(\bar{x},\bar{y})| P_X(du_1)P_X(du_2) < \infty.$$

Under the additional assumptions that the distributions of (X_0, X_n) and (X_n, X_0) have measure zero on the diagonal for all $n \in \mathbb{N}$ and that h as well as the eigenfunctions of (2.11) are bounded, he deduced the limit distribution of $n\,U_n$ using the spectral decomposition of h and invoking empirical process theory afterwards.

Dewan and Prakasa Rao [45] investigated stationary sequences of real-valued, associated random variables with summable covariances. Recall that $(X_n)_{n\in\mathbb{N}}$ is called associated if the inequalities $\operatorname{cov}(g_1(X_1,\dots,X_n), g_2(X_1,\dots,X_n)) \geq 0$ hold true for all $n \in \mathbb{N}$ and any coordinatewise nondecreasing functions $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2$, such that the above covariances exist. In order to derive the asymptotic distribution of a degenerate U-statistic, its kernel was assumed to be square integrable w.r.t. $P_X \times P_X$. The eigenvalues corresponding to (2.11) were supposed to be absolutely summable. Additionally, they assumed the related eigenfunctions to be differentiable and to satisfy $\sup_k \sup_x |\Phi_k^{(1)}(x)| < \infty$.

Under slightly modified assumptions, Huang and Zhang [74] established the analogous result for stationary, negatively associated sequences of centered, real-valued random variables, i.e. $\operatorname{cov}(g_1(X_i; i \in A), g_2(X_i; i \in B)) \leq 0$ holds true for all disjoint finite subsets $A, B \subset \mathbb{N}$ and coordinatewise nondecreasing real-valued functions g_1 and g_2. Suppose that U_n denotes a degenerate U-statistic with a kernel h satisfying $\iint h^2(x,y)P_X(dx)P_X(dy) < \infty$. As before, the related eigenvalues were supposed to be absolutely summable. Instead of a differentiability condition on the eigenfunctions, Huang and Zhang [74] assumed that these functions have bounded variation on any finite interval and $\sup_{k\geq 1} \mathbb{E}V^2_{\Phi_k}(X_1) < \infty$. Here, $V_{\Phi_k}(x)$ denotes the total variation of Φ_k on the interval $[0, x]$ if $x \geq 0$ and the total variation of Φ_k on the interval $[x, 0]$ if $x < 0$, respectively.

Sufficient conditions on h that assure the absolutely summable eigenvalues have already been discussed in Remark 2.2. We now illuminate the constraints concerning the eigenfunctions.

Remark 2.3. (i) Deriving conditions on the kernel functions that imply the above constraints regarding the eigenfunctions is intricate in general. Especially, the postulated uniformity in k of these constraints can hardly be proved without knowing the eigenfunctions explicitly. This in turn requires to solve the integral equation (2.11). De Wet and Venter [49] provide a method to derive the eigenfunctions associated with the kernels of (2.14). For distinct choices of the weight function w in these statistics, different types of eigenfunctions were obtained, e.g. suitably normalized Jacobi, Hermite and Laguerre polynomials. Thus, even within the class of statistics (2.14) the assumptions on the eigenfunctions have to be checked in a case-by-case manner. Furthermore, it is rather complicated or even impossible to solve the integral equation (2.11) for arbitrary h.

(ii) Eagleson [56] as well as Chen and White [25] did not impose conditions on the eigenfunctions. However, Borisov and Volodko [18] provide an example of a degenerate U-statistic for m-dependent random variables such that the prerequisites of both results are satisfied but the limit distribution does not coincide with the one of (2.13). This is due the fact that the spectral decomposition of the kernel does not remain valid for dependent observations in general. Additional conditions such as absolute continuity of the bivariate distributions of the underlying sample w.r.t. the corresponding product measure or regularity conditions on the kernel and the associated eigenfunctions are required. For instance, Borisov and Volodko [18] verified the validity of Eagleson's result under additional assumptions of that structure.

2.2.2 On empirical process approaches

Alternatively to the derivation of the asymptotic distributions of degenerate U- and V-type statistics via a spectral decomposition of the kernel, one can invoke empirical process theory. In case of real-valued i.i.d. random variables, corresponding results are presented in the book of Denker [44]. The idea is to express the statistics in terms of integrals w.r.t. the empirical process, i.e.

$$n\,V_n = n \iint_{\mathbb{R}\times\mathbb{R}} h(x,y) d(F_n - F)(x) d(F_n - F)(y).$$

The empirical distribution function of the underlying sample is denoted by F_n and the cumulative distribution function itself by F. The two main ingredients in order to obtain the asymptotic distribution of $n\,V_n$ are the approximation of the associated kernel functions by step functions and the convergence of the empirical process towards a Brownian bridge. More precisely, under the assumption $\iint_{\mathbb{R}^2} h^2(x,y)dF(x)dF(y) + \iint_{\mathbb{R}^2} h^2(x,x)dF(x)dF(x) < \infty$, one obtains

$$n\,V_n \xrightarrow{d} \iint_{(0,1)^2} h(F^{-1}(x), F^{-1}(y))\,dB(x)\,dB(y).$$

Here, B denotes the Brownian bridge on the interval [0,1]. If F is continuous, the corresponding U-statistic can be rewritten as

$$n\,U_n = \frac{n^2}{n-1} \iint_{\mathbb{R}\times\mathbb{R}} \bar{h}(x,y)d(F_n - F)(x)d(F_n - F)(y),$$

where $\bar{h}(x,y) = h(x,y)$ if $x \neq y$ and zero else. This is due the fact that the diagonal terms $h(X_i, X_i),\ i = 1, \ldots, n$, are not incorporated by U-statistics. Suppose that h is square integrable, then $n\,U_n \xrightarrow{d} \bar{U} := \iint_{(0,1)^2} \bar{h}(F^{-1}(x), F^{-1}(y))dB(x)dB(y)$. For sake of simplicity let $(X_k)_{k\in\mathbb{N}}$ be i.i.d. uniformly distributed on $[0,1]$. The connection to the previous subsection is then given by the relation

$$\bar{U} = \sum_{j=1}^{\infty} \lambda_j \left(\left[\int_{(0,1)} \Phi_j(x)dB(x) \right]^2 - 1 \right) \quad \text{a.s.},$$

where as before $(\lambda_k)_{k\in\mathbb{N}}$ and $(\Phi_k)_{k\in\mathbb{N}}$ denote eigenvalues and eigenfunctions associated with h.

This approach was adopted by Babbel [6] to develop weak invariance principles for U-statistics of stationary ϕ^*- and β-mixing sequences of random variables that are uniformly distributed on the interval $[0,1]$. The ϕ^*-mixing condition is defined similarly to the ϕ-mixing condition, but it additionally allows for interchanging the roles of $\mathcal{U}$ and $\mathcal{V}$ in the definition (2.1). Thus, the latter constraint is more restrictive than ϕ-mixing. We restrict ourselves to present the consequences of her results on ordinary degenerate V-statistics of stationary β-mixing sequences of random variables that are uniformly distributed on $[0,1]$. Suppose that

$$\begin{aligned} \iint_{(0,1)\times(0,1)} h(x,y)\,dP_{X_s,X_{s+m}}(x,y) &= 0, \quad \forall s \in \mathbb{Z},\ m \in \mathbb{N}, \qquad (2.15) \\ \mathbb{E}|n(n-1)U_n - (n-1)(n-2)U_{n-1}|^{2+\delta} &\leq C\,n^{(2+\delta)/2} \sup_{k\geq 0} \mathbb{E}|h(X_0, X_k)|^{2+\delta} \end{aligned}$$

for some $\delta > 0,\ C < \infty$, and $\beta_k = O(k^{-\theta})$ with $\theta = \max\{(2+\delta)(2+\gamma)/\delta, 5+\varepsilon\},\ \gamma, \varepsilon > 0$. Then,

$$n\,V_n \xrightarrow{d} \iint_{(0,1)^2} h(x,y)\,dK(1,x)\,dK(1,y).$$

Here, $(K(s,t))_{0\leq s,t\leq 1}$ denotes a Kiefer process with

$$\begin{aligned} \mathbb{E}K(t,s)K(\bar{t},\bar{s}) = \min\{t,\bar{t}\}\big\{&\mathbb{E}[(\mathbb{1}_{X_1\in(0,s]} - s)(\mathbb{1}_{X_1\in(0,\bar{s}]} - \bar{s})] \\ &+ \sum_{n=2}^{\infty} \mathbb{E}[(\mathbb{1}_{X_1\in(0,s]} - s)(\mathbb{1}_{X_n\in(0,\bar{s}]} - \bar{s}) + (\mathbb{1}_{X_n\in(0,s]} - s)(\mathbb{1}_{X_1\in(0,\bar{s}]} - \bar{s})]\big\}. \end{aligned}$$

The assumption concerning the decay of the mixing coefficients can be relaxed if the kernel is a bounded function. However, the condition (2.15) is rather restrictive. It turns out that it is in general not satisfied in the applications we consider in Chapter 5. Note that her method of using step functions in order to approximate the kernel by functions that are easier to continue work with can be interpreted as a Haar wavelet decomposition.

Remark 2.4. There are asymptotic results on degenerate U- and V-statistics under long-range dependence as well. For example, Dehling and Taqqu [41] derived their limit distributions by expressing the statistics by means of integrals similar as above and applying partial integration.

3 Asymptotic distributions of degenerate U- and V-statistics under τ-dependence

3.1 Motivation

Let $(X_n)_{n\in\mathbb{N}}$ be a stationary sequence of $\mathbb{R}^d$-valued random variables on some probability space $(\Omega, \mathcal{A}, P)$ with marginal distribution P_X. In this chapter we derive the limit distributions of

$$n\,U_n = \frac{1}{n-1}\sum_{\substack{j,k=1\\ j\neq k}}^{n} h(X_j, X_k) \qquad \text{and} \qquad n\,V_n = \frac{1}{n}\sum_{j,k=1}^{n} h(X_j, X_k),$$

where $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a measurable function that is symmetric in its arguments and satisfies the degeneracy condition $\int_{\mathbb{R}^d} h(x,y)P_X(dx) = 0,\ \forall\, y \in \mathbb{R}^d$. Throughout the remaining part of the thesis, the sequence $(X_n)_{n\in\mathbb{N}}$ is assumed to be τ-dependent in the sense of Definition 2.3. The objective is to establish asymptotic results under conditions that are easy to check.

Since the underlying concept of dependence is tailor-made for L_1-coupling techniques, it seems self-evident to investigate statistics with Lipschitz continuous kernels first. To use the method of proof presented in Subsection 2.2.1, i.e. applying a spectral decomposition of the function h, uniform Lipschitz continuity of the eigenfunctions $(\Phi_k)_{k\in\mathbb{N}}$ would be needed. However, the assumption of h being Lipschitz continuous merely implies

$$|\Phi_k(x_1) - \Phi_k(x_0)| = \frac{1}{|\lambda_k|}\left|\int_{\mathbb{R}^d} h(x_1,y) - h(x_0,y)\Phi_k(y)P_X(dy)\right| \leq \frac{L_h}{|\lambda_k|}\,|x_1 - x_0|\,\mathbb{E}\,|\Phi_k(X_1)|$$

if the corresponding eigenvalue λ_k is nonzero. We conjecture that the approach invoking a spectral decomposition of the kernel would again require conditions on the eigenfunctions. As it has been discussed in the previous section, it is difficult or even impossible to determine the eigenfunctions explicitly. Hence, we will not pursue this strategy of proof.

Still, we intend to develop a decomposition of the kernel that allows for the application of a central limit theorem. That is, the function h shall be approximated by a finite sum of functions which separate the two random variables that are accumulated in h. While Babbel [6] used a Haar wavelet decomposition for U-statistics of mixing processes, the application of Lipschitz continuous scale and wavelet functions is more suitable here in order to exploit the weak dependence property (2.5). In the first part of this chapter, we restrict ourselves to Lipschitz continuous kernels. In the subsequent section the problem of deriving the asymptotic distribution of U-statistics is reduced to the investigation of statistics with bounded kernels. Section 3.3 is dedicated to the wavelet decomposition of the kernel. Afterwards the limit distributions of the approximating statistics are derived and the asymptotics of $n\,U_n$ and $n\,V_n$ are deduced. Finally, we relax the smoothness assumptions concerning the kernel functions.

3.2 Approximation by statistics with bounded kernels

Using an appropriate kernel truncation, it is possible to approximate degenerate U-type statistics with unbounded kernels by statistics where the function h is bounded. To this end, let us assume:

(A1) **(i)** $(X_n)_{n\in\mathbb{N}}$ is a (strictly) stationary sequence of $\mathbb{R}^d$-valued random variables on some probability space $(\Omega, \mathcal{A}, P)$ with $\mathbb{E}\|X_1\|_1 < \infty$.

(ii) The sequence $(X_n)_{n\in\mathbb{N}}$ is τ-dependent in the sense of Definition 2.3 and the dependence coefficients $(\tau_r)_{r\in\mathbb{N}}$ satisfy $\sum_{r=1}^{\infty} r\,\tau_r^{\delta} < \infty$ for some $\delta \in (0,1)$.

Besides the conditions on the dependence structure of $(X_n)_{n\in\mathbb{N}}$, we make the following assumptions concerning the kernel:

(A2) **(i)** The kernel $h: \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}$ is a symmetric, measurable function and degenerate under P_X, i.e. $\int_{\mathbb{R}^d} h(x,y)P_X(dx) = 0,\ \forall\, y\in\mathbb{R}^d$.

(ii) For a δ satisfying (A1)(ii), the following moment constraints hold true with some $\nu > (2-\delta)/(1-\delta)$:

$$\sup_{k\in\mathbb{N}} \mathbb{E}|h(X_1, X_{1+k})|^{\nu} < \infty \quad \text{and} \quad \mathbb{E}\big|h(X_1, \widetilde{X}_1)\big|^{\nu} < \infty,$$

where $\widetilde{X}_1$ is an independent copy of X_1.

and

(A3) The kernel h is Lipschitz continuous.

Remark 3.1. The assumptions (A1) and (A3) imply the moment conditions of (A2)(ii) if additionally $\mathbb{E}\|X_1\|_1^{\nu} < \infty$. This can be easily verified by means of Minkowski's inequality.

Under these conditions we obtain the following result.

Lemma 3.1. *Suppose that the conditions (A1), (A2), and (A3) are fulfilled. Then there exists a family of bounded functions $(h_c)_{c\in\mathbb{R}_+}$, $h_c: \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}$, satisfying (A2) and (A3) uniformly and such that*

$$\lim_{c\to\infty} \limsup_{n\to\infty} n^2\,\mathbb{E}(U_n - U_{n,c})^2 = 0, \tag{3.1}$$

where $U_{n,c} = [n(n-1)]^{-1} \sum_{j,k=1,\,j\neq k}^{n} h_c(X_j, X_k)$.

Obviously, the difference $U_n - U_{n,c}$ is also a degenerate U-type statistic. Thus, Lemma 3.1 is an assertion on the second moment of a U-statistic. In future approximation steps those terms have to be investigated several times. Therefore, we insert a brief discussion on second moments of U-statistics under weak dependence before the proof of Lemma 3.1.

Lemma 3.2. *Let Z_n be a degenerate U-statistic with symmetric kernel H. Additionally, suppose that the assumptions (A1), (A2), and (A3) are satisfied. Then,*

$$\sup_{n\in\mathbb{N}} n^2\,\mathbb{E}Z_n^2 < \infty.$$

Remark 3.2. (i) This result immediately implies that the V-statistic associated with the kernel H also satisfies the above moment inequality if $\mathbb{E}H^2(X_1,X_1)<\infty$ since

$$n^2\,\mathbb{E}\left[\frac{1}{n^2}\sum_{j,k=1}^{n} H(X_j,X_k)\right]^2 \le 2\,n^2\,\mathbb{E}Z_n^2 + \frac{2}{n^2}\sum_{j,k=1}^{n}\mathbb{E}H(X_j,X_j)H(X_k,X_k).$$

(ii) Throughout all proofs of the thesis, C denotes a positive finite generic constant that may change its value even within a single calculation.

Proof. The second moment of Z_n can be estimated from above as follows:

$$\begin{aligned}\mathbb{E}[n\,Z_n]^2 &\le \frac{8}{(n-1)^2}\sum_{\substack{1\le i,j,k,l\le n\\ i<j;k<l;i\le k}} |\mathbb{E}H(X_i,X_j)H(X_k,X_l)| \\ &\le 8\sup_{1\le k<n}\mathbb{E}|H(X_1,X_{1+k})|^2 + \frac{8}{(n-1)^2}\sum_{r=1}^{n-1}\sum_{t=1}^{4} Z_{n,r}^{(t)},\end{aligned} \tag{3.2}$$

where

$$\begin{aligned}
Z_{n,r}^{(1)} &:= \sum_{\substack{1\le i<j;k<l;j\le l\le n\\ r:=\min\{j,k\}-i\ge l-\max\{j,k\}}} \left|\mathbb{E}H(X_i,X_j)H(X_k,X_l) - \mathbb{E}H(X_i,\widetilde{X}_j^{(r)})H(\widetilde{X}_k^{(r)},\widetilde{X}_l^{(r)})\right|,\\
Z_{n,r}^{(2)} &:= \sum_{\substack{1\le i<j;i\le k;k<l\le n\\ r:=l-\max\{j,k\}>\min\{j,k\}-i}} \left|\mathbb{E}H(X_i,X_j)H(X_k,X_l) - \mathbb{E}H(X_i,X_j)H(X_k,\widetilde{X}_l^{(r)})\right|,\\
Z_{n,r}^{(3)} &:= \sum_{\substack{1\le i\le k<l<j\le n\\ r:=k-i\ge j-l}} \left|\mathbb{E}H(X_i,X_j)H(X_k,X_l) - \mathbb{E}H(X_i,\widetilde{X}_j^{(r)})H(\widetilde{X}_k^{(r)},\widetilde{X}_l^{(r)})\right|,\\
Z_{n,r}^{(4)} &:= \sum_{\substack{1\le i\le k<l<j\le n\\ r:=j-l>k-i}} \left|\mathbb{E}H(X_i,X_j)H(X_k,X_l) - \mathbb{E}H(X_i,\widetilde{X}_j^{(r)})H(X_k,X_l)\right|.
\end{aligned}$$

Here, in every summand of $Z_{n,r}^{(1)}$ and $Z_{n,r}^{(3)}$ the vector $(\widetilde{X}_j^{(r)\prime},\widetilde{X}_k^{(r)\prime},\widetilde{X}_l^{(r)\prime})'$ is chosen such that it is independent of the random variable X_i, $(\widetilde{X}_j^{(r)\prime},\widetilde{X}_k^{(r)\prime},\widetilde{X}_l^{(r)\prime})' \stackrel{d}{=} (X_j',X_k',X_l')'$, and

$\mathbb{E}\|(\widetilde{X}_j^{(r)\prime}, \widetilde{X}_k^{(r)\prime}, \widetilde{X}_l^{(r)\prime})' - (X_j', X_k', X_l')'\|_1 \le \tau_r$. Within $Z_{n,r}^{(2)}$ (respectively $Z_{n,r}^{(4)}$), the random variable $\widetilde{X}_l^{(r)}$ (respectively $\widetilde{X}_j^{(r)}$) is chosen to be independent of the vector $(X_i', X_j', X_k')'$ (respectively $(X_i', X_k', X_l')'$) such that $\widetilde{X}_l^{(r)} \stackrel{d}{=} X_l$ (respectively $\widetilde{X}_j^{(r)} \stackrel{d}{=} X_j$) and $\mathbb{E}\|X_l - \widetilde{X}_l^{(r)}\|_1 \le \tau_r$ (respectively $\mathbb{E}\|X_j - \widetilde{X}_j^{(r)}\|_1 \le \tau_r$). This may possibly require an enlargement of the underlying probability space, see Lemma 2.1. Note that all subtrahends within the definition of $Z_{n,r}^{(t)},\ t = 1, \ldots, 4$, vanish according to the degeneracy of h. Moreover, it is important to point out that the number of summands of $Z_{n,r}^{(t)},\ t = 1, \ldots, 4$, is bounded by $(r+1)\, n^2$. For further calculations we restrict ourselves to $Z_{n,r}^{(1)}$. The remaining terms can be treated in an analogous manner. Every summand of $Z_{n,r}^{(1)}$ can be bounded by applying Hölder's inequality iteratively,

$$
\begin{aligned}
&\left|\mathbb{E}H(X_i,X_j)H(X_k,X_l) - \mathbb{E}H(X_i,\widetilde{X}_j^{(r)})H(\widetilde{X}_k^{(r)},\widetilde{X}_l^{(r)})\right| \\
&\le \left|\mathbb{E}H(X_k,X_l)\left[H(X_i,X_j) - H(X_i,\widetilde{X}_j^{(r)})\right]\right| \\
&\quad + \left|\mathbb{E}H(X_i,\widetilde{X}_j^{(r)})\left[H(X_k,X_l) - H(\widetilde{X}_k^{(r)},\widetilde{X}_l^{(r)})\right]\right| \\
&\le C\,\tau_r^\delta\left(\mathbb{E}\,|H(X_k,X_l)|^{1/(1-\delta)}\left|H(X_i,X_j) - H(X_i,\widetilde{X}_j^{(r)})\right|\right)^{1-\delta} \\
&\quad + C\,\tau_r^\delta\left(\mathbb{E}\left|H(X_i,\widetilde{X}_j^{(r)})\right|^{1/(1-\delta)}\left|H(X_k,X_l) - H(\widetilde{X}_k^{(r)},\widetilde{X}_l^{(r)})\right|\right)^{1-\delta} \\
&\le C\,\tau_r^\delta\left\{\left[\sup_{k\in\mathbb{N}}\mathbb{E}\,|H(X_1,X_{1+k})|^{(2-\delta)/(1-\delta)} + \mathbb{E}\left|H(X_1,\widetilde{X}_1)\right|^{(2-\delta)/(1-\delta)}\right]^{1/(2-\delta)}\right. \\
&\quad\times\left.\left[\sup_{k\in\mathbb{N}}\mathbb{E}\,|H(X_1,X_{1+k})|^{(2-\delta)/(1-\delta)} + \mathbb{E}\left|H(X_1,\widetilde{X}_1)\right|^{(2-\delta)/(1-\delta)}\right]^{(1-\delta)/(2-\delta)}\right\}^{1-\delta} \\
&\le C\,\tau_r^\delta.
\end{aligned}
\tag{3.3}
$$

Hence, we obtain $8\,(n-1)^{-2}\sum_{r=1}^n Z_{n,r}^{(1)} \le C\,(n-1)^{-2}\sum_{r=1}^n (r+1)\,n^2\,\tau_r^\delta \le C$. Moreover, $\sup_{1\le k<n}\mathbb{E}|H(X_1,X_{1+k})|^2$ is finite by (A2)(ii), which finally yields the assertion. □

These considerations facilitate the proof of Lemma 3.1.

Proof of Lemma 3.1. A kernel truncation argument is used to verify the assertion. For $c \in \mathbb{R}_+$ define $c_h := \max_{x,y\in[-c,c]^d}|h(x,y)|$ and set

$$
\widetilde{h}^{(c)}(x,y) := \begin{cases} h(x,y) & \text{for } |h(x,y)| \le c_h, \\ -c_h & \text{for } h(x,y) < -c_h, \\ c_h & \text{for } h(x,y) > c_h. \end{cases} \tag{3.4}
$$

Of course, the function $\widetilde{h}^{(c)}$ is not degenerate in general. This property has to be established artificially. To this end, we define the degenerate version of $\widetilde{h}^{(c)}$ by

$$
\begin{aligned}
h_c(x,y) :=\, &\widetilde{h}^{(c)}(x,y) - \int_{\mathbb{R}^d}\widetilde{h}^{(c)}(x,y)P_X(dx) - \int_{\mathbb{R}^d}\widetilde{h}^{(c)}(x,y)P_X(dy) \\
&+ \iint_{\mathbb{R}^d\times\mathbb{R}^d}\widetilde{h}^{(c)}(x,y)\,P_X(dx)\,P_X(dy).
\end{aligned}
$$

The approximation error $n^2\,\mathbb{E}(U_n - U_{n,c})^2$ can be reformulated in terms of $\mathbb{E}[n\,Z_n]^2$ with kernel $H = H^{(c)} := h - h_c$. Hence, in view of inequality (3.2), it remains to verify that $\sup_{k\in\mathbb{N}} \mathbb{E}|H^{(c)}(X_1, X_{1+k})|^2$ and $\limsup_{n\to\infty}(n-1)^{-2}\sum_{t=1}^{4}\sum_{r=1}^{n-1} Z_{n,r}^{(t)}$ tend to zero as $c\to\infty$. We investigate $\limsup_{n\to\infty}(n-1)^{-2}\sum_{r=1}^{n-1} Z_{n,r}^{(1)}$ only since the cases $t = 2,3,4$ can be treated in a similar manner. The summands of $Z_{n,r}^{(1)}$ can be bounded as follows:

$$
\begin{aligned}
&\left|\mathbb{E}H^{(c)}(X_i,X_j)H^{(c)}(X_k,X_l) - \mathbb{E}H^{(c)}(X_i,\widetilde{X}_j)H^{(c)}(\widetilde{X}_k,\widetilde{X}_l)\right| \\
&\le \mathbb{E}\left|H^{(c)}(X_k,X_l)\left[H^{(c)}(X_i,X_j) - H^{(c)}(X_i,\widetilde{X}_j)\right]\mathbb{1}_{(X_k',X_l')'\in[-c,c]^{2d}}\right| \\
&\quad + \mathbb{E}\left|H^{(c)}(X_k,X_l)\left[H^{(c)}(X_i,X_j) - H^{(c)}(X_i,\widetilde{X}_j)\right]\mathbb{1}_{(X_k',X_l')'\notin[-c,c]^{2d}}\right| \\
&\quad + \mathbb{E}\left|H^{(c)}(X_i,\widetilde{X}_j)\left[H^{(c)}(X_k,X_l) - H^{(c)}(\widetilde{X}_k,\widetilde{X}_l)\right]\mathbb{1}_{(X_i',\widetilde{X}_j')'\in[-c,c]^{2d}}\right| \\
&\quad + \mathbb{E}\left|H^{(c)}(X_i,\widetilde{X}_j)\left[H^{(c)}(X_k,X_l) - H^{(c)}(\widetilde{X}_k,\widetilde{X}_l)\right]\mathbb{1}_{(X_i',\widetilde{X}_j')'\notin[-c,c]^{2d}}\right| \\
&=: E_1 + E_2 + E_3 + E_4.
\end{aligned}
\tag{3.5}
$$

For sake of notational simplicity we suppress the upper indices r that appeared in the proof of Lemma 3.2. Note that $H^{(c)}$ is Lipschitz continuous with a constant that can be chosen independently of c. Therefore, an iterative application of Hölder's inequality to E_2 yields

$$
\begin{aligned}
E_2 \le& \left(\mathbb{E}\left|H^{(c)}(X_i,X_j) - H^{(c)}(X_i,\widetilde{X}_j)\right|\right)^{\delta} \\
&\times\left(\mathbb{E}\left|H^{(c)}(X_k,X_l)\right|^{1/(1-\delta)}\left|H^{(c)}(X_i,X_j) - H^{(c)}(X_i,\widetilde{X}_j)\right|\mathbb{1}_{(X_k',X_l')'\notin[-c,c]^{2d}}\right)^{1-\delta} \\
\le& C\,\tau_r^{\delta}\left\{\left(\mathbb{E}\left|H^{(c)}(X_k,X_l)\right|^{(2-\delta)/(1-\delta)}\mathbb{1}_{(X_k',X_l')'\notin[-c,c]^{2d}}\right)^{1/(2-\delta)}\right. \\
&\left.\left[\mathbb{E}\left|H^{(c)}(X_i,X_j)\right|^{(2-\delta)/(1-\delta)} + \mathbb{E}\left|H^{(c)}(X_i,\widetilde{X}_j)\right|^{(2-\delta)/(1-\delta)}\right]^{(1-\delta)/(2-\delta)}\right\}^{1-\delta} \\
\le& C\tau_r^{\delta}\left(\mathbb{E}\left|H^{(c)}(X_k,X_l)\right|^{(2-\delta)/(1-\delta)}\mathbb{1}_{(X_k',X_l')'\notin[-c,c]^{2d}}\right)^{(1-\delta)/(2-\delta)} \\
\le& C\,\tau_r^{\delta}\left[P(X_1\notin[-c,c]^d)\right]^{(1-\delta)/(2-\delta)-1/\nu}
\end{aligned}
\tag{3.6}
$$

according to (A2)(ii) and because of

$$
\begin{aligned}
&\mathbb{E}|H^{(c)}(X_i,X_j)|^{(2-\delta)/(1-\delta)} + \mathbb{E}|H^{(c)}(X_i,\widetilde{X}_j)|^{(2-\delta)/(1-\delta)} \\
&\le C\left(\mathbb{E}\sup_{k\in\mathbb{N}}|h(X_1,X_{1+k})|^{(2-\delta)/(1-\delta)} + \mathbb{E}|h(X_i,\widetilde{X}_j)|^{(2-\delta)/(1-\delta)}\right).
\end{aligned}
$$

Thus, we obtain $E_2 \le \tau_r^{\delta}\,\varepsilon_1(c)$ with $\varepsilon_1(c) \longrightarrow_{c\to\infty} 0$. Analogous calculations lead to $E_4 \le \tau_r^{\delta}\,\varepsilon_2(c)$ with $\varepsilon_2(c) \longrightarrow_{c\to\infty} 0$. Likewise, the approximation methods for E_1 and E_3

are equal. Therefore, only E_1 is investigated:

$$\begin{aligned}
E_1 \le\, & \mathbb{E}\left|\int_{\mathbb{R}^d} \widetilde{h}^{(c)}(X_k, y) P_X(dy) \left[H^{(c)}(X_i, X_j) - H^{(c)}(X_i, \widetilde{X}_j)\right] \mathbb{1}_{X_k \in [-c,c]^d}\right| \\
& + \mathbb{E}\left|\int_{\mathbb{R}^d} \widetilde{h}^{(c)}(y, X_l) P_X(dy) \left[H^{(c)}(X_i, X_j) - H^{(c)}(X_i, \widetilde{X}_j)\right] \mathbb{1}_{X_l \in [-c,c]^d}\right| \\
& + \mathbb{E}\left|\iint_{\mathbb{R}^d\times\mathbb{R}^d} \widetilde{h}^{(c)}(x, y) P_X(dx) P_X(dy) \left[H^{(c)}(X_i, X_j) - H^{(c)}(X_i, \widetilde{X}_j)\right]\right| \\
=\, & E_{1,1} + E_{1,2} + E_{1,3}.
\end{aligned}$$

Similarly to (3.6) we obtain

$$\begin{aligned}
E_{1,1} &\le C\,\tau_r^\delta \Big\{ \Big(\mathbb{E}\Big| \int_{\mathbb{R}^d} h(X_k, y) - \widetilde{h}^{(c)}(X_k, y) P_X(dy)\Big|^{(2-\delta)/(1-\delta)} \mathbb{1}_{X_k \in [-c,c]^d}\Big)^{1/(2-\delta)} \\
&\quad \times \Big[\sup_{k\in\mathbb{N}} \mathbb{E}\Big|H^{(c)}(X_1, X_{1+k})\Big|^{(2-\delta)/(1-\delta)} + \mathbb{E}\Big|H^{(c)}(X_i, \widetilde{X}_j)\Big|^{(2-\delta)/(1-\delta)}\Big]^{(1-\delta)/(2-\delta)}\Big\}^{1-\delta} \\
&\le C\,\tau_r^\delta \left(\iint_{\mathbb{R}^d\times\mathbb{R}^d} \left|h(x,y) - \widetilde{h}^{(c)}(x,y)\right|^{(2-\delta)/(1-\delta)} P_X(dy) \mathbb{1}_{x\in[-c,c]^d} P_X(dx)\right)^{(1-\delta)/(2-\delta)} \\
&\le \tau_r^\delta\, \varepsilon_3(c)
\end{aligned}$$

with $\varepsilon_3(c) \longrightarrow_{c\to\infty} 0$. The consideration of $E_{1,2}$ coincides with the previous one. The expression $E_{1,3}$ can be estimated as follows:

$$\begin{aligned}
E_{1,3} &\le C\,\tau_r \iint_{\mathbb{R}^d\times\mathbb{R}^d} \left|h(x,y) - \widetilde{h}^{(c)}(x,y)\right| P_X(dx)\, P_X(dy) \\
&\le C\,\tau_r \iint_{\mathbb{R}^d\times\mathbb{R}^d} |h(x,y)| \mathbb{1}_{(x',y')'\notin[-c,c]^{2d}}\, P_X(dx)\, P_X(dy) \\
&\le \tau_r\, \varepsilon_4(c)
\end{aligned}$$

with $\varepsilon_4(c) \longrightarrow_{c\to\infty} 0$. To sum up, we have $E_1 + E_2 + E_3 + E_4 \le \varepsilon_5(c)\,\tau_r^\delta$, where $\varepsilon_5(c) \longrightarrow_{c\to\infty} 0$ uniformly in n. This leads to

$$\lim_{c\to\infty} \limsup_{n\to\infty} \frac{1}{(n-1)^2} \sum_{r=1}^{n-1} Z_{n,r}^{(1)} \le \lim_{c\to\infty} \limsup_{n\to\infty} \frac{1}{(n-1)^2} \sum_{r=1}^{n-1} \left[(r+1)\, n^2\, \tau_r^\delta\, \varepsilon_5(c)\right] = 0.$$

In an analogous way, equivalent results can be established for $Z_{n,r}^{(i)},\ i = 2,3,4$. It remains to examine

$$\begin{aligned}
\sup_{k\in\mathbb{N}} \mathbb{E}[H^{(c)}(X_1, X_{1+k})]^2 &\le C\,\Big(\sup_{k\in\mathbb{N}} \mathbb{E}\big[h(X_1, X_{1+k}) - \widetilde{h}^{(c)}(X_1, X_{1+k})\big]^2 \\
&\qquad + \mathbb{E}\big[h(X_1, \widetilde{X}_1) - \widetilde{h}^{(c)}(X_1, \widetilde{X}_1)\big]^2\Big) \\
&\le C\big(\sup_{k\in\mathbb{N}} \mathbb{E}\, |h(X_1, X_{1+k})|^2\, \mathbb{1}_{(X_1', X_{1+k}')'\notin[-c,c]^{2d}} \\
&\qquad + \mathbb{E}\big|h(X_1, \widetilde{X}_1)\big|^2 \mathbb{1}_{(X_1', \widetilde{X}_1')'\notin[-c,c]^{2d}}\big).
\end{aligned}$$

Here, $\widetilde{X}_1$ denotes an independent copy of X_1. Applying Hölder's inequality once again, we obtain $\lim_{c\to\infty} \sup_{k\in\mathbb{N}} \mathbb{E}[H^{(c)}(X_1, X_{1+k})]^2 = 0$, which completes the proof. $\square$

Remark 3.3. It becomes apparent from the proof that any other truncation method can be employed instead of (3.4) as long as the following conditions are satisfied:

1. $\widetilde{h}^{(c)}(x,y) = h(x,y),\ \forall\, x,y \in [-c,c]$,
2. $\widetilde{h}^{(c)}$ is bounded for all $c \in \mathbb{R}_+$,
3. $(\widetilde{h}^{(c)})_c$ is Lipschitz continuous uniformly in c,
4. $\sup_{c\in\mathbb{R}_+} \sup_{k\in\mathbb{N}} \mathbb{E}|\widetilde{h}^{(c)}(X_1, X_{1+k})|^\nu < \infty$ and $\sup_{c\in\mathbb{R}_+} \mathbb{E}|\widetilde{h}^{(c)}(X_1, \widetilde{X}_1)|^\nu < \infty$.

3.3 A wavelet decomposition of the kernel

3.3.1 Some facts about wavelets

After the simplification of the problem in the previous section, we intend to develop a decomposition of the kernel such that a central limit theorem can be employed. Instead of imitating the proof from the i.i.d. case, a wavelet decomposition is invoked. As we have already broached in Section 3.1, Lipschitz continuity of the approximating functions is preferable according to the L_1-coupling property of the τ-dependence coefficient.

Let ϕ and ψ denote scale and wavelet functions associated with a one-dimensional multiresolution analysis. Since the pioneer work of Daubechies [28], it is well-known that these functions can be chosen in such a manner that they possess the following properties:

1. ϕ and ψ are real-valued and Lipschitz continuous,
2. ϕ and ψ have compact support,
3. $\int_{\mathbb{R}} \phi(x)\, dx = 1$ and $\int_{\mathbb{R}} \psi(x)\, dx = 0$.

An orthonormal basis of the space of square integrable functions $L_2(\mathbb{R}^d)$ can be constructed based on ϕ and ψ. For this purpose define $E := \{0,1\}^d \backslash \{0_d\}$, where 0_d denotes the d-dimensional null vector. In addition, set

$$\varphi^{(i)} := \begin{cases} \phi & \text{for } i = 0, \\ \psi & \text{for } i = 1 \end{cases}$$

and define functions $\Psi_{j,k}^{(e)} : \mathbb{R}^d \to \mathbb{R}$ for $j \in \mathbb{Z}, k = (k_1, \ldots, k_d)' \in \mathbb{Z}^d$, and $e = (e_1, \ldots, e_d)' \in E$ by

$$\Psi_{j,k}^{(e)}(x) := 2^{jd/2} \prod_{i=1}^{d} \varphi^{(e_i)}(2^j x_i - k_i), \quad x = (x_1, \ldots, x_d)' \in \mathbb{R}^d.$$

The system $(\Psi_{j,k}^{(e)})_{e\in E, j\in\mathbb{Z}, k\in\mathbb{Z}^d}$ is an orthonormal basis of $L_2(\mathbb{R}^d)$, see Wojtaszczyk [109], Section 5. The same holds true for $(\Phi_{j_0,k})_{k\in\mathbb{Z}^d} \bigcup (\Psi_{j,k}^{(e)})_{e\in E, j\geq j_0, k\in\mathbb{Z}^d}$, $j_0 \in \mathbb{Z}$, where $\Phi_{j,k}$:

$\mathbb{R}^d \to \mathbb{R}$ is given by $\Phi_{j,k}(x) := 2^{jd/2} \prod_{i=1}^d \phi(2^j x_i - k_i)$ for $j \in \mathbb{Z}, k \in \mathbb{Z}^d$. Therefore, every function $g : \mathbb{R}^d \to \mathbb{R}$ that belongs to $L_2(\mathbb{R}^d)$ can be represented as

$$g = \sum_{k \in \mathbb{Z}^d} \alpha_{j_0,k} \, \Phi_{j_0,k} + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}^d} \sum_{e \in E} \beta_{j,k}^{(e)} \, \Psi_{j,k}^{(e)}, \quad \forall \, j_0 \in \mathbb{Z}.$$

The involved coefficients are given by $\alpha_{j_0,k} = \int_{\mathbb{R}^d} g(u) \, \Phi_{j_0,k}(u) \, du$ and $\beta_{j,k}^{(e)} = \int_{\mathbb{R}^d} g(u) \, \Psi_{j,k}^{(e)}(u) \, du$, respectively.

We do not intend to summarize all important characteristics of the functions introduced above. Instead we concentrate on two auxiliary results that will be essential for a wavelet approximation of the kernel function h. Note that the functions involved in the both assertions below are not necessarily elements of the space $L_2(\mathbb{R}^d)$.

A first result is concerned with the question of how Lipschitz continuity of a function g is preserved in its approximation $g_j := \sum_{k \in \mathbb{Z}} \alpha_{j,k} \, \Phi_{j,k}$. It turns out that that the functions g_j are Lipschitz continuous uniformly in $j \in \mathbb{Z}$ although $\mathrm{Lip}(\Phi_{j,k}) = O(2^{j(d/2+1)})$.

Lemma 3.3. *Given a Lipschitz continuous function $g : \mathbb{R}^d \to \mathbb{R}$, define a wavelet approximation g_j by $g_j(x) := \sum_{k \in \mathbb{Z}^d} \alpha_{j,k} \, \Phi_{j,k}(x), j \in \mathbb{Z}$, with $\alpha_{j,k} = \int_{\mathbb{R}^d} g(u) \, \Phi_{j,k}(u) \, du$. Then the functions g_j are Lipschitz continuous with a constant that is independent of j.*

Proof. In order to establish Lipschitz continuity, the function g_j is decomposed into two parts:

$$\begin{aligned} g_j(x) &= \sum_{k \in \mathbb{Z}^d} \left[\int_{\mathbb{R}^d} \Phi_{j,k}(u) \, g(x) \, du \right] \Phi_{j,k}(x) + \sum_{k \in \mathbb{Z}^d} \left[\int_{\mathbb{R}^d} \Phi_{j,k}(u) [g(u) - g(x)] \, du \right] \Phi_{j,k}(x) \\ &=: H_1(x) + H_2(x). \end{aligned} \tag{3.7}$$

According to the above choice of the scale function (with characteristics 1. - 3.) the prerequisites of Corollary 8.1 of Härdle et al. [70] are fulfilled. That means that $\int_{\mathbb{R}} \sum_{l \in \mathbb{Z}} \phi(y - l) \phi(z - l) \, dz = 1, \; \forall \, y \in \mathbb{R}$. Based on this result, we obtain

$$\sum_{k \in \mathbb{Z}^d} \int_{\mathbb{R}^d} \Phi_{j,k}(u) \, \Phi_{j,k}(x) \, du = 2^{jd} \prod_{i=1}^d \int_{-\infty}^{\infty} \sum_{l \in \mathbb{Z}} \phi(2^j u_i - l) \, \phi(2^j x_i - l) \, du_i = 1, \quad \forall x \in \mathbb{R}^d,$$

by applying an appropriate variable substitution. To this end, note that for every fixed x the number of non-vanishing summands can be bounded by a finite constant uniformly in j because of the finite support of ϕ. Therefore, the order of summation and integration is interchangeable. Hence, $H_1 = g$ which in turn immediately implies the desired continuity property of H_1.

In order to analyse the function H_2, we define a sequence of functions $(\kappa_k)_{k \in \mathbb{Z}}$ by

$$\kappa_k(x) := \int_{\mathbb{R}^d} \Phi_{j,k}(u) \, [g(u) - g(x)] \, du.$$

These functions are Lipschitz continuous with a constant decreasing in j,

$$|\kappa_k(x) - \kappa_k(\bar{x})| \leq \mathrm{Lip}(g)\, O\left(2^{-jd/2}\right) \|x - \bar{x}\|_1, \tag{3.8}$$

since an appropriate variable substitution implies that $\int_{\mathbb{R}^d} |\Phi_{j,k}(u)|\, du = O(2^{-jd/2})$. Moreover, boundedness and Lipschitz continuity of ϕ yield

$$\|\Phi_{j,k}\|_\infty = O\left(2^{jd/2}\right) \quad \text{and} \quad |\Phi_{j,k}(x) - \Phi_{j,k}(\bar{x})| = O\left(2^{j(d/2+1)}\right) \|x - \bar{x}\|_1. \tag{3.9}$$

Thus,

$$\begin{aligned} |H_2(x) - H_2(\bar{x})| &\leq \sum_{k\in\mathbb{Z}^d} |\Phi_{j,k}(x)|\, |\kappa_k(x) - \kappa_k(\bar{x})| + \sum_{k\in\mathbb{Z}^d} |\kappa_k(\bar{x})|\, |\Phi_{j,k}(x) - \Phi_{j,k}(\bar{x})| \\ &\leq C\, \|x - \bar{x}\|_1 + \sum_{k\in\mathbb{Z}^d} |\kappa_k(\bar{x})|\, |\Phi_{j,k}(x) - \Phi_{j,k}(\bar{x})| \,. \end{aligned} \tag{3.10}$$

Now, two cases have to be distinguished in order to bound the second summand:

(1) $\bar{x} \in \mathrm{supp}\ (\Phi_{j,k})$,

(2) $\bar{x} \notin \mathrm{supp}\ (\Phi_{j,k})$.

(Here, supp denotes the support of a function.) In the first case, it is helpful to illuminate $|\kappa_k(\bar{x})| = |\int_{\mathbb{R}^d} \Phi_{j,k}(u)[g(u) - g(\bar{x})]\, du|$. The integrand is non-trivial only if $u \in \mathrm{supp}\ (\Phi_{j,k})$. For these values of u we obtain $|g(u) - g(\bar{x})| = O(2^{-j})$ due to Lipschitz continuity of g and the bounded support of ϕ. Consequently, we get

$$|\kappa_k(\bar{x})| \leq O\left(2^{-j}\right) \int_{\mathbb{R}^d} |\Phi_{j,k}(u)|\, du = O\left(2^{-j(d/2+1)}\right)$$

and $|\kappa_k(\bar{x})|\, |\Phi_{j,k}(x) - \Phi_{j,k}(\bar{x})| \leq C\, \|x - \bar{x}\|_1$. This in turn leads to

$$\sum_{k\in\mathbb{Z}^d} |\kappa_k(\bar{x})|\, |\Phi_{j,k}(x) - \Phi_{j,k}(\bar{x})| = C\, \|x - \bar{x}\|_1$$

as the number of non-vanishing summands can be bounded by a finite constant, independently of the values of x and $\bar{x}$. Therefore, Lipschitz continuity of H_2 is obtained as long as $\bar{x} \in \mathrm{supp}\ (\Phi_{j,k})$.

In the opposite case (2), we only have to consider the situation of $x \in \mathrm{supp}\ (\Phi_{j,k})$ since the setting $\bar{x}$, $x \notin \mathrm{supp}\ (\Phi_{j,k})$ is trivial. With the aid of (3.8) and (3.9), the first term of the right-hand side[1] of

$$|\kappa_k(\bar{x})\, [\Phi_{j,k}(x) - \Phi_{j,k}(\bar{x})]| \leq |\kappa_k(\bar{x}) - \kappa_k(x)|\, |\Phi_{j,k}(x)| + |\kappa_k(x)|\, |\Phi_{j,k}(x) - \Phi_{j,k}(\bar{x})| \tag{3.11}$$

can be estimated from above by $O\left(\|x - \bar{x}\|_1\right)$. The analysis of the second summand is identical to the analysis of the case $\bar{x} \in \mathrm{supp}\ (\Phi_{j,k})$.

Finally, we obtain $|H_2(x) - H_2(\bar{x})| \leq C\|x - \bar{x}\|_1$ where $C < \infty$ is a constant that is independent of j. □

[1] Throughout the rest of the thesis, *r.h.s.* abbreviates the expression 'right-hand side'.

While the previous lemma is dedicated to the smoothness of approximating functions, the next assertion addresses uniform approximation properties on compact intervals.

Lemma 3.4. *Let $g : \mathbb{R}^d \to \mathbb{R}$ be a function that is continuous on some interval $(-c, c)^d$. For arbitrary $b \in (0, c)$ and $L \in \mathbb{N}$, there exists a $J_0(L, b, c) \in \mathbb{N}$ such that for g and its approximation g_j, given by $g_j(x) = \sum_{k\in\mathbb{Z}^d} \alpha_{j,k}\Phi_{j,k}(x)$, it holds the inequality*

$$\max_{x\in[-b,b]^d} |g(x) - g_j(x)| \le 1/L, \quad \forall\, j \ge J_0(L, b, c).$$

Proof. Given $b \in (0, c)$, we define a function $\bar{g}^{(b,c)} : \mathbb{R}^d \to \mathbb{R}$ by $\bar{g}^{(b,c)}(x) := g(x)\, w_{b,c}(x)$, where $w_{b,c}$ is a Lipschitz continuous and nonnegative weight function with compact support $S_w \subset (-c, c)^d$. Moreover, $w_{b,c}$ is assumed to be bounded from above by one and $w_{b,c}(x) := 1$ for $x \in (-b - \delta, b + \delta)^d$ for some $\delta > 0$ with $b + \delta < c$. Additionally, we set $\alpha_{j,k}^{(b,c)} := \int_{\mathbb{R}^d} \bar{g}^{(b,c)}(u)\, \Phi_{j,k}(u)\, du$. Hence, we obtain

$$\begin{aligned}
&\max_{x\in[-b,b]^d} |g(x) - g_j(x)| \\
&\le \max_{x\in[-b,b]^d} \left| \bar{g}^{(b,c)}(x) - \sum_{k\in\mathbb{Z}^d} \alpha_{j,k}^{(b,c)} \Phi_{j,k}(x) \right| + \max_{x\in[-b,b]^d} \left| \sum_{k\in\mathbb{Z}^d} \alpha_{j,k}^{(b,c)} \Phi_{j,k}(x) - g_j(x) \right| \\
&=: \max_{x\in[-b,b]^d} A^{(j)}(x) + \max_{x\in[-b,b]^d} B^{(j)}(x).
\end{aligned}$$

Since $\bar{g}^{(b,c)} \in C_0(\mathbb{R}^d)$, Theorem 8.4 of Wojtaszczyk [109] implies

$$\max_{x\in[-b,b]^d} A^{(j)}(x) \underset{j\to\infty}{\longrightarrow} 0.$$

Thus, there exists a $\bar{J}_0(K, b, c)$ such that $\max_{x\in[-b,b]^d} A^{(j)}(x) \le 1/L$ for all $j \ge \bar{J}_0(K, b, c)$. Moreover, the introduction of the finite set of indices

$$\bar{Z}(j) := \left\{ k \in \mathbb{Z}^d \,|\, \Phi_{j,k}(x) \ne 0 \text{ for some } x \in [-b, b]^d \right\}$$

leads to

$$\max_{x\in[-b,b]^d} B^{(j)}(x) = \max_{x\in[-b,b]^d} \left| \sum_{k\in\bar{Z}(j)} \left(\alpha_{j,k} - \alpha_{j,k}^{(b,c)} \right) \Phi_{j,k}(x) \right|.$$

This term is equal to zero if the length of the support of ϕ is bounded by $2^j\, \delta$ since the definition of $\bar{g}^{(b,c)}$ then implies $\alpha_{j,k} = \alpha_{j,k}^{(b,c)}$, $\forall\, k \in \bar{Z}(j)$. This finally yields the assertion with some $J_0(L, b, c) \ge \bar{J}_0(L, b, c)$. □

3.3.2 Further approximation steps: Using wavelet decompositions

When deriving the asymptotic distributions of degenerate U- and V-type statistics, the application of a central limit theorem is hampered by the fact that the kernel h aggregates the involved random variables in a possibly complicated manner. In order to separate the variables, we approximate the function h with the aid of multi-dimensional wavelets.

An L_2-approximation of the statistic $U_{n,c}$ by a statistic based on a wavelet approximation of the bounded kernel h_c can be established. To this end, we introduce functions $\widetilde{h}_c^{(K,L)}$ with

$$\begin{aligned}\widetilde{h}_c^{(K,L)}(x,y) := & \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{k_1,k_2}^{(c)}\, \Phi_{0,k_1}(x)\, \Phi_{0,k_2}(y) \\ & + \sum_{j=0}^{J(L)-1} \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \sum_{e\in\bar{E}} \beta_{j;k_1,k_2}^{(c,e)}\, \Psi_{j;k_1,k_2}^{(e)}(x,y),\end{aligned} \tag{3.12}$$

where $\bar{E} := (E\times E)\cup(E\times\{0_d\})\cup(\{0_d\}\times E)$,

$$\Psi_{j;k_1,k_2}^{(e)} := \begin{cases} \Psi_{j,k_1}^{(e_1)}\, \Psi_{j,k_2}^{(e_2)} & \text{for } (e_1',e_2')'\in E\times E, \\ \Psi_{j,k_1}^{(e_1)}\, \Phi_{j,k_2} & \text{for } (e_1',e_2')'\in E\times\{0_d\}, \\ \Phi_{j,k_1}\, \Psi_{j,k_2}^{(e_2)} & \text{for } (e_1',e_2')'\in \{0_d\}\times E, \end{cases}$$

$$\alpha_{k_1,k_2}^{(c)} := \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c(x,y)\, \Phi_{0,k_1}(x)\, \Phi_{0,k_2}(y)\, dx\, dy,$$

and

$$\beta_{j;k_1,k_2}^{(c,e)} := \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c(x,y)\, \Psi_{j;k_1,k_2}^{(e)}(x,y)\, dx\, dy.$$

We refer to the degenerate version of $\widetilde{h}_c^{(K,L)}$ as $h_c^{(K,L)}$, given by

$$\begin{aligned}h_c^{(K,L)}(x,y) := \widetilde{h}_c^{(K,L)}(x,y) - \int_{\mathbb{R}^d} \widetilde{h}_c^{(K,L)}(x,y) P_X(dx) - \int_{\mathbb{R}^d} \widetilde{h}_c^{(K,L)}(x,y) P_X(dy) \\ + \iint_{\mathbb{R}^d\times\mathbb{R}^d} \widetilde{h}_c^{(K,L)}(x,y) P_X(dx) P_X(dy).\end{aligned}$$

The associated U-type statistic will be denoted by $U_{n,c}^{(K,L)}$. The subsequent assertion assures that the approximation error is asymptotically negligible.

Lemma 3.5. *Assume that (A1), (A2), and (A3) are fulfilled. Then, the sequence $(J(L))_{L\in\mathbb{N}}$ in (3.12) with $J(L) \longrightarrow_{L\to\infty} \infty$ can be chosen such that*

$$\lim_{L\to\infty} \lim_{K\to\infty} \limsup_{n\to\infty} n^2\, \mathbb{E}\left(U_{n,c} - U_{n,c}^{(K,L)}\right)^2 = 0.$$

Proof. The assertion is verified in two steps. First, the bounded kernel h_c, constructed in the proof of Lemma 3.1, is approximated by $\widetilde{h}_c^{(L)}$ which is defined by

$$\widetilde{h}_c^{(L)}(x,y) := \sum_{k_1,k_2\in\mathbb{Z}^d} \alpha_{J(L);k_1,k_2}^{(c)}\, \Phi_{J(L),k_1}(x)\, \Phi_{J(L),k_2}(y)$$

with $\alpha_{J(L);k_1,k_2}^{(c)} := \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c(x,y)\, \Phi_{J(L),k_1}(x)\, \Phi_{J(L),k_2}(y)\, dx\, dy$. Here, the indices $(J(L))_{L\in\mathbb{N}}$ with $J(L) \longrightarrow_{L\to\infty} \infty$ are chosen such that the assertion of Lemma 3.4 holds true for $c = 2b$ and $b = b(L) \in \mathbb{R}$ such that $P(X_1 \notin [-b(L), b(L)]^d) \le 1/L$. As the functions $\widetilde{h}_c^{(L)}$, $L \in \mathbb{N}$, are not degenerate in general, we introduce their degenerate counterparts $h_c^{(L)}$, given by

$$\begin{aligned} h_c^{(L)}(x,y) := &\widetilde{h}_c^{(L)}(x,y) - \int_{\mathbb{R}^d} \widetilde{h}_c^{(L)}(x,y) P_X(dx) - \int_{\mathbb{R}^d} \widetilde{h}_c^{(L)}(x,y) P_X(dy) \\ &+ \iint_{\mathbb{R}^d \times \mathbb{R}^d} \widetilde{h}_c^{(L)}(x,y) P_X(dx) P_X(dy), \end{aligned}$$

and denote the corresponding U-statistics by $U_{n,c}^{(L)}$.
Now, the structure of the proof is as follows: First, we prove

$$\limsup_{n\to\infty} n^2\, \mathbb{E}\left(U_{n,c} - U_{n,c}^{(L)}\right)^2 \underset{L\to\infty}{\longrightarrow} 0. \tag{3.13}$$

Thus, the main goal of the first step is the multiplicative separation of the random variables which are cumulated in h_c. The aim of the second step is the approximation of $h_c^{(L)}$, whose representation is given by an infinite sum, by a function consisting of only finitely many summands. That is, it remains to show that for every fixed L

$$\limsup_{n\to\infty} n^2\, \mathbb{E}\left(U_{n,c}^{(L)} - U_{n,c}^{(K,L)}\right)^2 \underset{K\to\infty}{\longrightarrow} 0. \tag{3.14}$$

<u>*Step 1:*</u> $\limsup_{n\to\infty} n^2\, \mathbb{E}(U_{n,c} - U_{n,c}^{(L)})^2 \longrightarrow_{L\to\infty} 0$.
In order to verify (3.13), we rewrite $n^2\, \mathbb{E}(U_{n,c} - U_{n,c}^{(L)})^2$ in terms of $\mathbb{E}[n\, Z_n]^2$ with kernel function $H = H^{(L)} := h_c - h_c^{(L)}$. Hence, it remains to verify that $\limsup_{n\to\infty}(n-1)^{-2}\sum_{r=1}^{n-1}\sum_{t=1}^{4} Z_{n,r}^{(t)}$, defined in the proof of Lemma 3.2, and $\sup_{k\in\mathbb{N}} \mathbb{E}|H^{(L)}(H_1, X_{1+k})|^2$ tend to zero as $L \to \infty$. Exemplarily, we consider the case $t = 1$. The summands of $Z_{n,r}^{(1)}$ can be bounded as follows:

$$\begin{aligned} &\left|\mathbb{E}H^{(L)}(X_i,X_j)H^{(L)}(X_k,X_l) - H^{(L)}(X_i,\widetilde{X}_j)H^{(L)}(\widetilde{X}_k,\widetilde{X}_l)\right| \\ &\le \mathbb{E}\left|H^{(L)}(X_k,X_l)\left[H^{(L)}(X_i,X_j) - H^{(L)}(X_i,\widetilde{X}_j)\right]\right| \\ &\quad + \mathbb{E}\left|H^{(L)}(X_i,\widetilde{X}_j)\left[H^{(L)}(X_k,X_l) - H^{(L)}(\widetilde{X}_k,\widetilde{X}_l)\right]\right|. \end{aligned}$$

Since further approximations are similar for both summands, we concentrate on the first one. Note that boundedness of h_c implies uniform boundedness of $(H^{(L)})_L$ due to the compact support of the function ϕ. Moreover, the constant $\mathrm{Lip}(H^{(L)})$ does not depend on L in consequence of Lemma 3.3. Therefore, the application of Hölder's inequality leads to

$$\mathbb{E}\left|H^{(L)}(X_k,X_l)\left[H^{(L)}(X_i,X_j) - H^{(L)}(X_i,\widetilde{X}_j)\right]\right| \le C\,\tau_r^\delta \left[\mathbb{E}|H^{(L)}(X_k,X_l)|^{1/(1-\delta)}\right]^{1-\delta}.$$

In order to analyse $\mathbb{E}|H^{(L)}(X_k,X_l)|^{1/(1-\delta)}$, recall that the sequence $(b(L))_L$ is chosen such that $P(X_1 \notin [-b(L), b(L)]^d) \le 1/L$. This allows for the following approximation:

$$\begin{aligned} &\mathbb{E}|H^{(L)}(X_k,X_l)|^{1/(1-\delta)} \\ &= \mathbb{E}|H^{(L)}(X_k,X_l)|^{1/(1-\delta)}\mathbb{1}_{X_k,X_l\in[-b(L),b(L)]^d} + O\left(P(X_1 \notin [-b(L),b(L)]^d)\right) \\ &\le \max_{x,y\in[-b(L),b(L)]^d} |H^{(L)}(x,y)|^{1/(1-\delta)} + \frac{C}{L}. \end{aligned}$$

According to Lemma 3.4 and the above choice of the sequence $(b(L))_L$, we obtain

$$\begin{aligned}
&\max_{x,y\in[-b(L),b(L)]^d} \left|H^{(L)}(x,y)\right| \\
&\leq \frac{1}{L} + 2 \max_{x,y\in[-b(L),b(L)]^d} \mathbb{E}\left|h_c(x,X_1) - \widetilde{h}_c^{(L)}(x,X_1)\right| \\
&\quad + \left|\iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c(x,y) - \widetilde{h}_c^{(L)}(x,y)\, P_X(dx)\, P_X(dy)\right| \\
&\leq \frac{4}{L} + 2 \max_{x\in[-b(L),b(L)]^d} \mathbb{E}\left|h_c(x,X_1) - \widetilde{h}_c^{(L)}(x,X_1)\right| \mathbb{1}_{X_1\notin[-b(L),b(L)]^d} \\
&\quad + 2\int_{\mathbb{R}^d}\int_{\mathbb{R}^d\setminus[-b(L),b(L)]^d} \left|h_c(x,y) - \widetilde{h}_c^{(L)}(x,y)\right| P_X(dx)\, P_X(dy) \\
&\leq \frac{C}{L}.
\end{aligned}$$

Consequently,

$$\left|\mathbb{E}H^{(L)}(X_i,X_j)H^{(L)}(X_k,X_l) - \mathbb{E}H^{(L)}(X_i,\widetilde{X}_j)H^{(L)}(\widetilde{X}_k,\widetilde{X}_l)\right| \leq C\,\varepsilon_L\,\tau_r^{\delta},$$

where $(\varepsilon_L)_L$ is a certain null sequence. This implies that $\limsup_{n\to\infty} Z_{n,r}^{(1)}$ tends to zero as L increases. Furthermore, one obtains $\sup_{k\in\mathbb{N}} \mathbb{E}[H^{(L)}(X_1,X_{1+k})]^2 = O(1/L)$ similarly to the consideration of $\mathbb{E}|H^{(L)}(X_k,X_l)|^{1/(1-\delta)}$ above. Thus, we get $\limsup_{n\to\infty} n^2\,\mathbb{E}(U_{n,c} - U_{n,c}^{(L)})^2 \longrightarrow_{L\to\infty} 0$.

Step 2: $\limsup_{n\to\infty} n^2\,\mathbb{E}(U_{n,c}^{(L)} - U_{n,c}^{(K,L)})^2 \longrightarrow_{K\to\infty} 0$.

Before starting with the actual approximations, we collect various properties of the involved functions that will be invoked in the calculations thereafter. Since the scale function ϕ and the wavelet ψ have compact support, the number of overlapping functions within $(\Phi_{0,k})_{k\in\{-K,\dots,K\}^d}$ and $(\Psi_{j,k}^{(e)})_{k\in\{-K,\dots,K\}^d, 0\leq j<J(L), e\in E}$ can be bounded by a constant that is independent of K. This leads to uniform Lipschitz continuity of $(h_c^{(K,L)})_{K\in\mathbb{N}}$ by Lipschitz continuity of the functions ϕ and ψ. Moreover, note that the sequence $(H^{(K)})_K$ with $H^{(K)} := h_c^{(L)} - h_c^{(K,L)}$ is uniformly bounded. The function $\widetilde{h}_c^{(L)}$ can be represented as follows:

$$\widetilde{h}_c^{(L)}(x,y) = \sum_{k_1,k_2\in\mathbb{Z}^d} \alpha_{k_1,k_2}^{(c)}\, \Phi_{0,k_1}(x)\, \Phi_{0,k_2}(y) + \sum_{j=0}^{J(L)-1} \sum_{k_1,k_2\in\mathbb{Z}^d} \sum_{e\in\bar{E}} \beta_{j;k_1,k_2}^{(c,e)}\, \Psi_{j;k_1,k_2}^{(e)}(x,y).$$

This equivalence results from the properties of multiresolution analyses.

Analogously to step 1 of the proof, the approximation error $n^2\,\mathbb{E}(U_{n,c}^{(L)} - U_{n,c}^{(K,L)})^2$ is reformulated in terms of $\mathbb{E}[n\,Z_n]^2$ with kernel $H = H^{(K)}$. As before, we exemplarily take $(n-1)^{-2}\sum_{r=1}^{n-1} Z_{n,r}^{(1)}$ and $\sup_{k\in\mathbb{N}} \mathbb{E}|H^{(K)}(X_1,X_{1+k})|^2$ into further consideration. The summands of $Z_{n,r}^{(1)}$ can be approximated by

$$\begin{aligned}
&\left|\mathbb{E}H^{(K)}(X_i,X_j)H^{(K)}(X_k,X_l)-\mathbb{E}H^{(K)}(X_i,\widetilde{X}_j)H^{(K)}(\widetilde{X}_k,\widetilde{X}_l)\right|\\
&\le \mathbb{E}\left|H^{(K)}(X_k,X_l)\left[H^{(K)}(X_i,X_j)-H^{(K)}(X_i,\widetilde{X}_j)\right]\mathbb{1}_{(X_k',X_l')'\in[-B,B]^{2d}}\right|\\
&\quad+\mathbb{E}\left|H^{(K)}(X_k,X_l)\left[H^{(K)}(X_i,X_j)-H^{(K)}(X_i,\widetilde{X}_j)\right]\mathbb{1}_{(X_k',X_l')'\notin[-B,B]^{2d}}\right|\\
&\quad+\mathbb{E}\left|H^{(K)}(X_i,\widetilde{X}_j)\left[H^{(K)}(X_k,X_l)-H^{(K)}(\widetilde{X}_k,\widetilde{X}_l)\right]\mathbb{1}_{(X_i',\widetilde{X}_j')'\in[-B,B]^{2d}}\right|\\
&\quad+\mathbb{E}\left|H^{(K)}(X_i,\widetilde{X}_j)\left[H^{(K)}(X_k,X_l)-H^{(K)}(\widetilde{X}_k,\widetilde{X}_l)\right]\mathbb{1}_{(X_i',\widetilde{X}_j')'\notin[-B,B]^{2d}}\right|\\
&=: E_1+E_2+E_3+E_4
\end{aligned}$$

for arbitrary $B>0$. Obviously, it suffices to illuminate the first two summands. The both remaining terms can be treated in a similar manner. Due to the preliminary considerations at the beginning of this step of proof, one can choose $(B=B(K,L))_{K\in\mathbb{N}}$ such that $\max_{x,y\in[-B,B]^d}|\widetilde{h}_c^{(L)}(x,y)-\widetilde{h}_c^{(K,L)}(x,y)|=0$ and $B\longrightarrow_{K\to\infty}\infty$. This setting allows for the following estimations:

$$\begin{aligned}
E_1&\le C\,\tau_r^{\delta}\left[\mathbb{E}|H^{(K)}(X_k,X_l)|^{1/(1-\delta)}\,\mathbb{1}_{(X_k',X_l')'\in[-B,B]^d}\right]^{1-\delta}\\
&\le C\,\tau_r^{\delta}\Bigg[\mathbb{E}\int_{\mathbb{R}^d}|\widetilde{h}^{(L)}(y,X_1)-\widetilde{h}^{(K,L)}(y,X_1)|^{1/(1-\delta)}\,P_X(dy)\,\mathbb{1}_{X_1\notin[-B,B]^d}\\
&\qquad+\iint_{\mathbb{R}^d\times\mathbb{R}^d}|\widetilde{h}^{(L)}(x,y)-\widetilde{h}^{(K,L)}(x,y)|^{1/(1-\delta)}\,\mathbb{1}_{(x',y')'\notin[-B,B]^{2d}}\,P_X(dx)\,P_X(dy)\Bigg]^{1-\delta}\\
&\le C\,\tau_r^{\delta}\left[P(X_1\notin[-B,B]^d)\right]^{1-\delta}
\end{aligned}$$

and $E_2\le C\,\tau_r^{\delta}\left[P(X_1\notin[-B,B]^d)\right]^{1-\delta}$ according to uniform boundedness of the involved functions. Analogously, it can be shown that $\sup_{k\in\mathbb{N}}\mathbb{E}[H^{(K)}(X_1,X_{1+k})]^2\le CP(X_1\notin[-B,B]^d)$. Finally, we obtain

$$\limsup_{n\to\infty} n^2\,\mathbb{E}\left(U_{n,c}^{(L)}-U_{n,c}^{(K,L)}\right)^2\le C\,P(X_1\notin[-B(K),B(K)]^d)^{1-\delta}\left[\limsup_{n\to\infty}\sum_{r=1}^{n-1}(r+1)\,\tau_r^{\delta}\right],$$

where the r.h.s. decreases to zero as K tends to infinity. Hence, the assertion (3.14) holds. □

3.4 Derivation of the limit distributions

Based on the previous approximation steps, the limit distributions of degenerate U- and V-statistics are derived. First, a central limit theorem and the continuous mapping theorem are employed to obtain the asymptotic distribution of $n\,U_{n,c}^{(K,L)}$. To this end, the introduction of functions

$$\Psi_{j,k}^{(0_d)}:=\Phi_{j,k},\quad j\in\mathbb{Z},\ k\in\mathbb{Z}^d,$$

is useful.

Lemma 3.6. *Suppose that the conditions (A1), (A2), and (A3) are fulfilled. Then, as* $n \to \infty$,

$$n\, U_{n,c}^{(K,L)} \xrightarrow{d} Z_c^{(K,L)}$$

with

$$\begin{aligned} Z_c^{(K,L)} := & \sum_{k_1,k_2 \in \{-K,\dots,K\}^d} \alpha_{k_1,k_2}^{(c)} \left[Z_{k_1} Z_{k_2} - A_{k_1,k_2} \right] \\ & + \sum_{j=0}^{J(L)-1} \sum_{k_1,k_2 \in \{-K,\dots,K\}^d} \sum_{e=(e_1',e_2')' \in \bar{E}} \beta_{j;k_1,k_2}^{(c,e)} \left[Z_{j;k_1}^{(e_1)} Z_{j;k_2}^{(e_2)} - B_{j;k_1,k_2}^{(e)} \right]. \end{aligned} \tag{3.15}$$

Here, $A_{k_1,k_2} := \operatorname{cov}\left(\Phi_{0,k_1}(X_1)\,,\, \Phi_{0,k_2}(X_1)\right)$, $B_{j;k_1,k_2}^{(e)} := \operatorname{cov}(\Psi_{j,k_1}^{(e_1)}(X_1), \Psi_{j,k_2}^{(e_2)}(X_1))$, $(Z_k)_{k \in \{-K,\dots,K\}^d}$ *and* $(Z_{j;k}^{(e_1)})_{1 \le j < J(L); k \in \{-K,\dots,K\}^d; e_1 \in \{0,1\}^d}$ *are centered and jointly normally distributed random vectors. Their covariance structure is given by*

$$\begin{aligned} \operatorname{cov}(Z_{k_1}, Z_{k_2}) &= \operatorname{cov}(\Phi_{0,k_1}(X_1),\, \Phi_{0,k_2}(X_1)) \\ &\quad + \sum_{s=2}^{\infty} \left[\operatorname{cov}(\Phi_{0,k_1}(X_1),\, \Phi_{0,k_2}(X_s)) + \operatorname{cov}(\Phi_{0,k_1}(X_s),\, \Phi_{0,k_2}(X_1)) \right], \\ \operatorname{cov}(Z_{k_1}, Z_{j;k_2}^{(e_1)}) &= \operatorname{cov}(\Phi_{0,k_1}(X_1),\, \Psi_{j,k_2}^{(e_1)}(X_1)) \\ &\quad + \sum_{s=2}^{\infty} \left[\operatorname{cov}(\Phi_{0,k_1}(X_1),\, \Psi_{j,k_2}^{(e_1)}(X_s)) + \operatorname{cov}(\Phi_{0,k_1}(X_s),\, \Psi_{j,k_2}^{(e_1)}(X_1)) \right], \\ \operatorname{cov}(Z_{j_1;k_1}^{(e_1)}, Z_{j_2;k_2}^{(e_2)}) &= \operatorname{cov}(\Psi_{j_1,k_1}^{(e_1)}(X_1)\,,\, \Psi_{j_2,k_2}^{(e_2)}(X_1)) \\ &\quad + \sum_{s=2}^{\infty} \left[\operatorname{cov}(\Psi_{j_1,k_1}^{(e_1)}(X_1),\, \Psi_{j_2,k_2}^{(e_2)}(X_s)) + \operatorname{cov}(\Psi_{j_1,k_1}^{(e_1)}(X_s),\, \Psi_{j_2,k_2}^{(e_2)}(X_1)) \right] \end{aligned}$$

for $k_1, k_2 \in \{-K,\dots,K\}^d$, $j, j_1, j_2 \in \{1,\dots,J(L)-1\}$ *and* $(e_1', e_2')' \in \bar{E}$.

Proof. The following modified representation of $\widetilde{h}_c^{(K,L)}$ will be useful in the sequel:

$$\begin{aligned} \widetilde{h}_c^{(K,L)}(x,y) &= \sum_{k_1,k_2 \in \{-K,\dots,K\}^d} \alpha_{k_1,k_2}^{(c)}\, \Phi_{0,k_1}(x)\, \Phi_{0,k_2}(y) \\ &\quad + \sum_{j=0}^{J(L)-1} \sum_{k_1,k_2 \in \{-K,\dots,K\}^d} \sum_{(e_1',e_2')' \in \bar{E}} \beta_{j;k_1,k_2}^{(c,e)}\, \Psi_{j,k_1}^{(e_1)}(x)\, \Psi_{j,k_2}^{(e_2)}(y) \\ &= \sum_{k,l=1}^{M(K,L)} \gamma_{k,l}^{(c)}\, \tilde{q}_k(x)\, \tilde{q}_l(y), \end{aligned}$$

where $(\tilde{q}_l)_{l=1}^{M(K,L)}$ is an ordering of $\bigcup_{k \in \{-K,\dots,K\}^d} \left\{ \{\Phi_{0,k}\} \cup \{\Psi_{j,k}^{(e)}\}_{e \in \{0,1\}^d, j \in \{0,\dots,J(L)-1\}} \right\}$ and $\gamma_{k,l}^{(c)} = \gamma_{l,k}^{(c)}$, $k,l \in \{1,\dots,M(K,L)\}$, are the associated coefficients. Moreover, the introduction of $q_k(X_i) := \tilde{q}_k(X_i) - \mathbb{E}\tilde{q}_k(X_i)$, $k \in \{1,\dots,M(K,L)\}$, $i \in \{1,\dots,n\}$, allows for the compact notation of the statistic $n\,U_{n,c}^{(K,L)}$,

$$\begin{aligned} n\,U_{n,c}^{(K,L)} &= \frac{1}{n-1}\sum_{i\neq j} h_c^{(K,L)}(X_i,X_j) \\ &= \frac{n}{n-1}\sum_{k,l=1}^{M(K,L)} \gamma_{k,l}^{(c)} \left(\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} q_k(X_i)\right]\left[\frac{1}{\sqrt{n}}\sum_{j=1}^{n} q_l(X_j)\right] - \frac{1}{n}\sum_{i=1}^{n} q_k(X_i)q_l(X_i)\right). \end{aligned}$$

The latter summand in the round brackets converges to $\mathbb{E}q_k(X_1)\,q_l(X_1)$ in probability by the weak law of large numbers (Lemma 2.3). In order to derive the limit distribution of the first summand, we consider $n^{-1/2}\sum_{i=1}^{n}(q_1(X_i),...,q_{M(K,L)}(X_i))'$ first. Due to the Cramér-Wold device, it suffices to analyse $\sum_{k=1}^{M(K,L)} t_k\, n^{-1/2}\sum_{i=1}^{n} q_k(X_i)$, $\forall\,(t_1,...,t_{M(K,L)})' \in \mathbb{R}^{M(K,L)}$. Asymptotic normality can be established by applying Lemma 2.2 to $n^{-1/2}\sum_{i=1}^{n} Q_i$ with $Q_i := \sum_{k=1}^{M(K,L)} t_k\, q_k(X_i)$, $i = 1,\dots,n$. To this end, the prerequisites of that tool have to be checked. Obviously, we are given a strictly stationary sequence of centered and bounded random variables. According to the minimal L_1-coupling property of the τ-dependence coefficients we obtain

$$\tau\left(\sigma(Q_{s_1},\dots,Q_{s_u}),(Q_{t_1},Q_{t_2},Q_{t_3})'\right) \leq \mathbb{E}\left\|(Q_{t_1},Q_{t_2},Q_{t_3})' - (\widetilde{Q}_{t_1},\widetilde{Q}_{t_2},\widetilde{Q}_{t_3})'\right\|_1$$

for positive integers $s_1 \leq \cdots \leq s_u < s_u + r \leq t_1 \leq t_2 \leq t_3$. Here, the random variables $\widetilde{Q}_i$, $i = t_1, t_2, t_3$, are given by $\sum_{k=1}^{M(K,L)} t_k q_k(\widetilde{X}_i)$ and $(\widetilde{X}_{t_1},\widetilde{X}_{t_2},\widetilde{X}_{t_3})'$ is chosen such that the assertion of Lemma 2.1 holds true with $\mathcal{M} = \sigma(X_{s_1},\dots,X_{s_u})$ and $X = (X'_{t_1},X'_{t_2},X'_{t_3})'$. Therefore, the sequence $(Q_i)_i$ is τ-dependent with coefficients $\bar{\tau}_r$, $r \in \mathbb{N}$, that are bounded by $\bar{\tau}_r \leq \mathrm{Lip}(\sum_{k=1}^{M(K,L)} t_k q_k)\,\tau_r$ and thus summable.

To show that $n^{-1}\operatorname{var}(Q_1+\cdots+Q_n) \longrightarrow_{n\to\infty} \sigma^2 := \operatorname{var}(Q_1) + 2\sum_{r=2}^{\infty}\operatorname{cov}(Q_1,Q_r)$, the validity of assumption (A1) can be employed. Note that $\sigma^2 \leq \operatorname{var}(Q_1) + C\sum_{r=1}^{\infty}\bar{\tau}_r < \infty$ due to the boundedness of the function $\sum_{k=1}^{M(K,L)} t_k q_k$. We have

$$\begin{aligned} \left|\frac{1}{n}\operatorname{var}(Q_1+...+Q_n) - \sigma^2\right| &= \left|\frac{2}{n}\sum_{r=2}^{n}(n-[r-1])\operatorname{cov}(Q_1,Q_r) - 2\sum_{r=2}^{\infty}\operatorname{cov}(Q_1,Q_r)\right| \\ &\leq 2\sum_{r=2}^{\infty}\min\left\{\frac{r-1}{n},1\right\}\,|\operatorname{cov}(Q_1,Q_r)| \\ &\leq C\sum_{r=2}^{\infty}\min\left\{\frac{r-1}{n},1\right\}\tau_{r-1}. \end{aligned}$$

The summability of the dependence coefficients in connection with Lebesgue's dominated convergence theorem yields the desired result. Thus, all prerequisites of Lemma 2.2 are fulfilled and we obtain

$$n^{-1/2}(Q_1+...+Q_n) \xrightarrow{d} N(0,\sigma^2).$$

Eventually, the continuous mapping theorem and back-transformation of the representation of $h_c^{(K,L)}$ lead to

$$n\,U_{n,c}^{(K,L)} \xrightarrow{d} Z_c^{(K,L)} = \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{k_1,k_2}^{(c)} [Z_{k_1} Z_{k_2} - A_{k_1,k_2}] + \sum_{j=0}^{J(L)-1} \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \sum_{e=(e_1',e_2')'\in\bar{E}} \beta_{j;k_1,k_2}^{(c,e)} \left[Z_{j;k_1}^{(e_1)} Z_{j;k_2}^{(e_2)} - B_{j;k_1,k_2}^{(e)}\right].$$

The covariance structure of the involved centered normally distributed random variables results immediately from the foregoing considerations. □

In conjunction with Lemma 3.1 and Lemma 3.5, the previous result is applied to deduce the limit distribution of $n\,U_n$. Finally, the weak law of large numbers provided in Lemma 2.3 yields the asymptotics of $n\,V_n$ since $n\,V_n = [(n-1)/n]\,[n\,U_n + n^{-1}\sum_{k=1}^{n} h(X_k, X_k)]$.

Theorem 3.1. *Suppose that the assumptions (A1), (A2), and (A3) are fulfilled. Then, as $n \to \infty$,*

$$n\,U_n \xrightarrow{d} Z$$

with

$$Z := \lim_{c\to\infty} \Big(\sum_{k_1,k_2\in\mathbb{Z}^d} \alpha_{k_1,k_2}^{(c)} \,[Z_{k_1}\, Z_{k_2} - A_{k_1,k_2}] + \sum_{j=0}^{\infty} \sum_{k_1,k_2\in\mathbb{Z}^d} \sum_{e=(e_1',e_2')'\in\bar{E}} \beta_{j;k_1,k_2}^{(c,e)} \,[Z_{j;k_1}^{(e_1)}\, Z_{j;k_2}^{(e_2)} - B_{j;k_1,k_2}^{(e)}]\Big),$$

where the r.h.s. converges in the L_2-sense. The constants $((A_{k,l}, B_{j;k,l}^{(e)})')_{k,l\in\mathbb{Z}^d, j\in\mathbb{N}_0, e\in\bar{E}}$ are defined as in Lemma 3.6, $(Z_k)_{k\in\mathbb{Z}^d}$ and $(Z_{j;k}^{(e)})_{j\in\mathbb{N}_0,\, k\in\mathbb{Z}^d,\, e\in\{0,1\}^d}$ are centered and jointly normally distributed random vectors with a covariance structure as given in Lemma 3.6. If additionally $\mathbb{E}|h(X_1,X_1)| < \infty$, then, as $n \to \infty$,

$$n\,V_n \xrightarrow{d} Z + \mathbb{E}h(X_1, X_1).$$

Proof. Lemma 3.1 and Lemma 3.5 yield

$$\lim_{c\to\infty} \lim_{L\to\infty} \lim_{K\to\infty} \limsup_{n\to\infty} n^2\, \mathbb{E}(U_{n,c}^{(K,L)} - U_n)^2 = 0.$$

Moreover, $n\,U_{n,c}^{(K,L)} \xrightarrow{d} Z_c^{(K,L)}$ by virtue of Lemma 3.6. Hence, it remains to show that $\lim_{c\to\infty} \lim_{L\to\infty} \lim_{K\to\infty} \mathbb{E}(Z_c^{(K,L)} - Z)^2 = 0$ according to Billingsley [16], Theorem 4.2. To this end, we first show that $(Z_c^{(K,L)})_K$ is a Cauchy sequence in L_2. Analogously to Lemma 3.6, one proves that $n\,(U_{n,c}^{(K_1,L)} - U_{n,c}^{(K_2,L)}) \xrightarrow{d} Z_c^{(K_1,L)} - Z_c^{(K_2,L)}$. Now we employ the fact that $\mathbb{E}|Y| \leq \liminf_{n\to\infty} \mathbb{E}|Y_n|$ if $Y_n \xrightarrow{d} Y$, see for instance Billingsley [16], Theorem 5.3. In conjunction with the result (3.14), this allows for

$$\begin{aligned}
&\mathbb{E}\left(Z_c^{(K_1,L)} - Z_c^{(K_2,L)}\right)^2 \\
&\leq \liminf_{n\to\infty} n^2\,\mathbb{E}\left(U_{n,c}^{(K_1,L)} - U_{n,c}^{(K_2,L)}\right)^2 \\
&\leq 2\,\limsup_{n\to\infty} n^2\,\mathbb{E}\left(U_{n,c}^{(K_1,L)} - U_{n,c}^{(L)}\right)^2 + 2\,\limsup_{n\to\infty} n^2\,\mathbb{E}\left(U_{n,c}^{(K_2,L)} - U_{n,c}^{(L)}\right)^2 \underset{K_1,K_2\to\infty}{\longrightarrow} 0.
\end{aligned}$$

Due to the completeness of the L_2, $n\,U_{n,c}^{(L)} \xrightarrow{d} Z_c^{(L)}$ with

$$\begin{aligned}
Z_c^{(L)} :=& \sum_{k_1,k_2\in\mathbb{Z}^d} \alpha_{k_1,k_2}^{(c)}\,[Z_{k_1}Z_{k_2} - A_{k_1,k_2}] \\
&+ \sum_{j=0}^{J(L)-1} \sum_{k_1,k_2\in\mathbb{Z}^d} \sum_{e=(e_1',e_2')'\in\bar{E}} \beta_{j;k_1,k_2}^{(c,e)}\,[Z_{j;k_1}^{(e_1)}\,Z_{j;k_2}^{(e_2)} - B_{j;k_1,k_2}^{(e)}],
\end{aligned}$$

where the sums converge in the L_2-sense. Note that interchanging of the finite summation and taking limits can be easily justified by applying step 2 of the proof of Lemma 3.5 and the previous considerations of this proof to each summand ($j = 0, \ldots, J(L)-1$) separately.

In a next step we have to prove that $n\,U_{n,c} \xrightarrow{d} Z_c$, where

$$\begin{aligned}
Z_c :=& \sum_{k_1,k_2\in\mathbb{Z}^d} \alpha_{k_1,k_2}^{(c)}[Z_{k_1}Z_{k_2} - A_{k_1,k_2}] \\
&+ \sum_{j=0}^{\infty} \sum_{k_1,k_2\in\mathbb{Z}^d} \sum_{e=(e_1',e_2')'\in\bar{E}} \beta_{j;k_1,k_2}^{(c,e)}[Z_{j;k_1}^{(e_1)}Z_{j;k_2}^{(e_2)} - B_{j;k_1,k_2}^{(e)}].
\end{aligned}$$

With similar arguments as before, the claim holds true since

$$\begin{aligned}
\mathbb{E}\left(Z_c^{(L_1)} - Z_c^{(L_2)}\right)^2 &\leq 4\limsup_{K\to\infty} \mathbb{E}\left(Z_c^{(K,L_1)} - Z_c^{(K,L_2)}\right)^2 \\
&\leq 4\limsup_{K\to\infty}\,\liminf_{n\to\infty} n^2\,\mathbb{E}\left(U_{n,c}^{(K,L_1)} - U_{n,c}^{(K,L_2)}\right)^2 \\
&\leq 16\,\limsup_{K\to\infty}\liminf_{n\to\infty} n^2\left[\mathbb{E}\left(U_{n,c}^{(K,L_1)} - U_{n,c}^{(L_1)}\right)^2 + \mathbb{E}\left(U_{n,c}^{(L_1)} - U_{n,c}^{(L_2)}\right)^2\right. \\
&\quad \left. + \mathbb{E}\left(U_{n,c}^{(K,L_2)} - U_{n,c}^{(L_2)}\right)^2\right] \\
&\leq 16\,\liminf_{n\to\infty} n^2\,\mathbb{E}\left(U_{n,c}^{(L_1)} - U_{n,c}^{(L_2)}\right)^2 \underset{L_1,L_2\to\infty}{\longrightarrow} 0
\end{aligned} \tag{3.16}$$

according to the result of convergence (3.13). Applying this method once again, we get $\lim_{c\to\infty} \mathbb{E}(Z - Z_c)^2 = 0$ by means of Lemma 3.1. Eventually, the relation $\lim_{c\to\infty}\lim_{L\to\infty}\lim_{K\to\infty} \mathbb{E}(Z_c^{(K,L)} - Z)^2 = 0$ follows from the previous considerations, which in turn leads to the desired limit distribution of $n\,U_n$.

Based on the result concerning U-type statistics, the limit distributions of $n\,V_n$ can be established. Since $V_n = [(n-1)/n]\,U_n + n^{-2}\sum_{i=1}^n h(X_i, X_i)$, it remains to verify that

$n^{-1}\sum_{i=1}^{n} h(X_i, X_i) \xrightarrow{P} \mathbb{E}h(X_1, X_1)$. This in turn is a consequence of Lemma 2.3. Thus, the proof is complete. □

Conclusion.

- As in the case of i.i.d. random variables, the asymptotic distributions of U- and V-statistics are basically weighted sums of products of centered normal random variables. In contrast to many other results in the literature, the prerequisites of the present theorem, namely moment constraints and Lipschitz continuity of the kernel, can be checked fairly easily in many cases. Note that the distribution of the limit variable does not depend on the specific choice of the scale and wavelet functions due to the uniqueness of the weak limit.
- Nevertheless, the asymptotic distribution still has a complicated structure. Hence, quantiles can hardly be determined on the basis of the previous result. However, this problem plays a minor role since we show in the next chapter that the conditional distributions of the bootstrap counterparts of $n\,U_n$ and $n\,V_n$, given $X_1, \ldots, X_n$, converge to the same limits in probability. Therefore, certain bootstrap algorithms can be employed to approximate quantiles of degenerate U- and V-type statistics.

3.5 Weakening the smoothness assumptions

Of course, the assumption of Lipschitz continuous kernels is quite restrictive and restrains the applicability of our results. As it becomes apparent in Section 5.3, the Lipschitz condition is violated even in cases of very simple parametric test statistics. There, an L_2-type test statistic based on the empirical characteristic function is investigated. This function is bounded and Lipschitz continuous. However, as soon as parameters have to be estimated, additional terms enter the corresponding V-type statistic. In general, these expressions are no longer bounded which destroys the Lipschitz continuity of the associated kernel due to the product structure of kernel functions that result from test statistics of L_2-type.

Therefore, we extend the asymptotic theory for U- and V-statistics to a broader class of kernel functions. Weakening the smoothness assumption goes along with additional moment constraints. Moreover, a faster decay of the dependence coefficients is required. Besides (A1) and (A2), we assume:

(A4) **(i)** The kernel function satisfies

$$|h(x,y) - h(\bar{x},\bar{y})| \leq f(x,\bar{x},y,\bar{y})\,[\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1]\,, \quad \forall\, x,\, \bar{x},\, y,\, \bar{y} \in \mathbb{R}^d,$$

for a continuous function $f : \mathbb{R}^{4d} \to \mathbb{R}$ that is symmetric under permutations of its arguments. Moreover, let $\delta \in (0,1)$ be such that (A2)(ii) holds and $\eta := 1/(1-\delta)$. Then,

$$\sup_{k_1,\ldots,k_5 \in \mathbb{N}} \mathbb{E}\left(\max_{a_1,a_2 \in [-A,A]^d} [f(Y_{k_1}, Y_{k_2}+a_1, Y_{k_3}, Y_{k_4}+a_2)]^\eta\, \|Y_{k_5}\|_1 \right) < \infty$$

for some $A > 0$ and any $(Y'_{k_1}, \ldots, Y'_{k_5})'$ consisting of independent subvectors $(Y'_{k_{j_1(m)}}, \ldots, Y'_{k_{j_l(m)}})' \stackrel{d}{=} (X'_{k_{j_1(m)}}, \ldots, X'_{k_{j_l(m)}})'$, $l, m = 1, \ldots, 5$.

(ii) The dependence coefficients satisfy $\sum_{r=1}^{\infty} r\, \tau_r^{\delta^2} < \infty$.

Obviously all functions satisfying assumption (A4)(i) constitute a subclass of functions that are Lipschitz continuous on any bounded interval. Even though the condition above has a rather technical structure, it is satisfied e.g. by polynomial kernel functions as long as the sample variables have sufficiently many finite moments.

As in Section 3.2 we start with an auxiliary result regarding the second moments of U-statistics with kernels satisfying the weaker regularity condition (A4).

Lemma 3.7. *Let Z_n be a degenerate U-statistic with symmetric kernel H. Additionally, suppose that the assumptions (A1), (A2), and (A4) are satisfied. Then,*

$$\sup_{n \in \mathbb{N}} n^2\, \mathbb{E} Z_n^2 < \infty.$$

Remark 3.4. The result remains valid for the corresponding V-statistic if additionally $\mathbb{E} H^2(X_1, X_1) < \infty$.

Proof. We follow the lines of the proof of Lemma 3.2. Merely, the estimation (3.3) has to be modified as follows:

$$\begin{aligned}
&\left|\mathbb{E} H(X_i, X_j) H(X_k, X_l) - \mathbb{E} H(X_i, \widetilde{X}_j^{(r)}) H(\widetilde{X}_k^{(r)}, \widetilde{X}_l^{(r)})\right| \\
&\leq \left(\mathbb{E}\, |H(X_k, X_l)|^{1/(1-\delta)} \left|H(X_i, X_j) - H(X_i, \widetilde{X}_j^{(r)})\right|\right)^{1-\delta} \\
&\qquad \times \left(\mathbb{E} f(X_i, X_j, X_i, \widetilde{X}_j^{(r)}) \|X_j - \widetilde{X}_j^{(r)}\|_1\right)^{\delta} \\
&\quad + \left(\mathbb{E} \left|H(X_i, \widetilde{X}_j^{(r)})\right|^{1/(1-\delta)} \left|H(X_k, X_l) - H(\widetilde{X}_k^{(r)}, \widetilde{X}_l^{(r)})\right|\right)^{1-\delta} \\
&\qquad \times \left(\mathbb{E} f(X_k, X_l, \widetilde{X}_k^{(r)}, \widetilde{X}_l^{(r)}) \left[\|X_k - \widetilde{X}_k^{(r)}\|_1 + \|X_l - \widetilde{X}_l^{(r)}\|_1\right]\right)^{\delta} \\
&\leq C\, \tau_r^{\delta^2} \left[\mathbb{E}\left(\left[f(X_i, X_j, X_i, \widetilde{X}_j^{(r)})\right]^{\eta} \left[\|X_j\|_1 + \|\widetilde{X}_j^{(r)}\|_1\right]\right)\right]^{\delta(1-\delta)} \\
&\quad + C\, \tau_r^{\delta^2} \left[\mathbb{E}\left(\left[f(X_k, X_l, \widetilde{X}_k^{(r)}, \widetilde{X}_l^{(r)})\right]^{\eta} \left[\|X_k\|_1 + \|\widetilde{X}_k^{(r)}\|_1 + \|X_l\|_1 + \|\widetilde{X}_l^{(r)}\|_1\right]\right)\right]^{\delta(1-\delta)} \\
&\leq C\, \tau_r^{\delta^2}.
\end{aligned}$$

□

Analogously to Lemma 3.1 and Lemma 3.5, the following assertion holds.

Lemma 3.8. *Suppose that (A1), (A2), and (A4) are fulfilled. Then, a family of bounded kernels $(h_c)_c$ satisfying (A2) and (A4) uniformly and the sequence of indices $(J(L))_{L \in \mathbb{N}}$ in (3.12) with $J(L) \longrightarrow_{L \to \infty} \infty$ can be chosen such that*

$$\lim_{c \to \infty} \lim_{L \to \infty} \lim_{K \to \infty} \limsup_{n \to \infty} n^2\, \mathbb{E}\left(U_n - U_{n,c}^{(K,L)}\right)^2 = 0.$$

Proof. In order to prove this result, we follow the lines of the proofs of Lemma 3.1, Lemma 3.3, as well as Lemma 3.5 and carry out some modifications.
Step 1: $\lim_{c\to\infty}\limsup_{n\to\infty} n^2\,\mathbb{E}(U_n - U_{n,c})^2 = 0.$
To reduce the problem to statistics with bounded kernels, we use the same truncation method as in the proof of Lemma 3.1 and the modified approximation

$$\left|H^{(c)}(x,y) - H^{(c)}(\bar{x},\bar{y})\right| \le [2\,f(x,\bar{x},y,\bar{y}) + g(x,\bar{x}) + g(y,\bar{y})]\,[\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1],$$

where the kernel $H^{(c)}$ is defined as in the proof of Lemma 3.1 and the function g is defined through $g(x,\bar{x}) := \int_{\mathbb{R}^d} f(x,\bar{x},z,z)P_X(dz)$. Obviously, the function f_1 given by $f_1(x,\bar{x},y,\bar{y}) := 2\,f(x,\bar{x},y,\bar{y}) + g(x,\bar{x}) + g(y,\bar{y})$ inherits the symmetry and moment properties of f. Under (A4)(i) Hölder's inequality yields

$$\begin{aligned}
&\mathbb{E}|H^{(c)}(Y_{k_1},Y_{k_2}) - H^{(c)}(Y_{k_3},Y_{k_4})| \\
&\le \left(\mathbb{E}[f_1(Y_{k_1},Y_{k_2},Y_{k_3},Y_{k_4})]^{1/(1-\delta)}\left[\sum_{i=1}^4 \|Y_{k_i}\|_1\right]\right)^{1-\delta} (\mathbb{E}\|Y_{k_1}-Y_{k_3}\|_1 + \mathbb{E}\|Y_{k_2}-Y_{k_4}\|_1)^\delta \\
&\le C\,(\mathbb{E}\|Y_{k_1}-Y_{k_3}\|_1 + \mathbb{E}\|Y_{k_2}-Y_{k_4}\|_1)^\delta
\end{aligned}$$

for Y_{k_i} $(k_i \in \mathbb{N},\ i = 1,\ldots,4)$ as defined in (A4). Plugging in this inequality into the calculations of the proof of Lemma 3.1 yields $\limsup_{n\to\infty} n^2\,\mathbb{E}(U_n - U_{n,c})^2 \longrightarrow_{c\to\infty} 0$.
Step 2: $\lim_{L\to\infty}\limsup_{n\to\infty} n^2\,\mathbb{E}(U_{n,c} - U_{n,c}^{(L)})^2 = 0.$
Since the kernel h_c of $U_{n,c}$ is no longer Lipschitz continuous, we cannot invoke Lemma 3.3 for the wavelet approximation of the bounded kernel. Therefore, the continuity properties of the kernel $\widetilde{h}_c^{(L)} := \sum_{k_1,k_2\in\mathbb{Z}^d} \alpha^{(c)}_{J(L);k_1,k_2}\Phi_{J(L),k_1}\Phi_{J(L),k_2}$ have to be elaborated at this point. Similarly to (3.7) and (3.10) in the proof of Lemma 3.3, we obtain

$$\begin{aligned}
&\left|\widetilde{h}_c^{(L)}(\bar{x},\bar{y}) - \widetilde{h}_c^{(L)}(x,y)\right| \\
&\le f_1(x,\bar{x},y,\bar{y})[\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1] + |H_2(\bar{x},\bar{y}) - H_2(x,y)| \\
&\le Cf_1(x,\bar{x},y,\bar{y})[\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1] \\
&\quad + \sum_{k_1,k_2\in\mathbb{Z}^d}\big(|\kappa_{k_1,k_2}(\bar{x},\bar{y})||\Phi_{J(L),k_1}(x)\Phi_{J(L),k_2}(y) - \Phi_{J(L),k_1}(\bar{x})\Phi_{J(L),k_2}(\bar{y})|\big),
\end{aligned} \tag{3.17}$$

where κ_{k_1,k_2} is given by $\kappa_{k_1,k_2}(x,y) := \iint_{\mathbb{R}^d\times\mathbb{R}^d}\Phi_{J(L),k_1}(u)\Phi_{J(L),k_2}(v)[h_c(u,v) - h_c(x,y)]\,du\,dv$ and the function H_2 is defined as in the proof of Lemma 3.3. In order to approximate the last summand of (3.17), the cases whether or not $(\bar{x}',\bar{y}')' \in \operatorname{supp}\left(\Phi_{J(L),k_1}\Phi_{J(L),k_2}\right)$ are distinguished. In the first case, we achieve an upper bound of order

$$O\left(\max_{a_1,a_2\in[-S_\phi/2^{J(L)},S_\phi/2^{J(L)}]^d} f_1(\bar{x},\bar{x}+a_1,\bar{y},\bar{y}+a_2)(\|\bar{x}-x\|_1 + \|\bar{y}-y\|_1)\right)$$

invoking analogous arguments as in the proof of Lemma 3.3 since

$$\begin{aligned}
|\kappa_{k_1,k_2}(\bar{x},\bar{y})| &\leq \frac{S_\phi}{2^{J(L)}} \max_{a_1,a_2\in[-S_\phi/2^{J(L)},S_\phi/2^{J(L)}]^d} f_1(\bar{x},\bar{x}+a_1,\bar{y},\bar{y}+a_2) \\
&\quad \times \iint_{\mathbb{R}^d\times\mathbb{R}^d} |\Phi_{J(L),k_1}(u)\Phi_{J(L),k_2}(v)|\,du\,dv \\
&\leq O\left(2^{-J(L)(d+1)}\right) \max_{a_1,a_2\in[-S_\phi/2^{J(L)},S_\phi/2^{J(L)}]^d} f_1(\bar{x},\bar{x}+a_1,\bar{y},\bar{y}+a_2).
\end{aligned}$$

Here, S_ϕ denotes the length of the support of ϕ. In the second case, a decomposition similar to (3.11) can be employed which leads to the upper bound

$$O\left(f_1(x,\bar{x},y,\bar{y}) + \max_{a_1,a_2\in[-S_\phi/2^{J(L)},S_\phi/2^{J(L)}]^d} f_1(x,x+a_1,y,y+a_2)\right)(\|\bar{x}-x\|_1+\|\bar{y}-y\|_1).$$

Consequently, we obtain

$$\begin{aligned}
&\left|\widetilde{h}_c^{(L)}(\bar{x},\bar{y}) - \widetilde{h}_c^{(L)}(x,y)\right| \\
&\leq O\Big(f_1(x,\bar{x},y,\bar{y}) + \max_{a_1,a_2\in[-S_\phi/2^{J(L)},S_\phi/2^{J(L)}]^d} f_1(x,x+a_1,y,y+a_2) \\
&\quad + \max_{a_1,a_2\in[-S_\phi/2^{J(L)},S_\phi/2^{J(L)}]^d} f_1(\bar{x},\bar{x}+a_1,\bar{y},\bar{y}+a_2)\Big)(\|\bar{x}-x\|_1+\|\bar{y}-y\|_1) \\
&=: f_2(x,\bar{x},y,\bar{y})(\|\bar{x}-x\|_1+\|\bar{y}-y\|_1).
\end{aligned}$$

This leads to $|H^{(L)}(x,y) - H^{(L)}(\bar{x},\bar{y})| \leq f_3(x,\bar{x},y,\bar{y})(\|x-\bar{x}\|_1+\|y-\bar{y}\|_1)$ with $f_3(x,\bar{x},y,\bar{y}) := 2\,f_2(x,\bar{x},y,\bar{y}) + \int_{\mathbb{R}^d} f_2(x,\bar{x},z,z)P_X(dz) + \int_{\mathbb{R}^d} f_2(z,z,y,\bar{y})P_X(dz)$, where the kernels $H^{(L)}$ are defined as in step 1 of the proof of Lemma 3.5. Note that under (A4)(i) we have $\mathbb{E}[f_3(Y_i,Y_j,Y_k,Y_l)]^\eta\,(\|Y_i\|_1+\|Y_j\|_1+\|Y_k\|_1+\|Y_l\|_1) < \infty$ if $J(L)$ is sufficiently large. Thus, one gets

$$\mathbb{E}|H^{(L)}(Y_{k_1},Y_{k_2}) - H^{(L)}(Y_{k_3},Y_{k_4})| \leq C(\mathbb{E}\|Y_{k_1}-Y_{k_3}\|_1 + \mathbb{E}\|Y_{k_2}-Y_{k_4}\|_1)^\delta \tag{3.18}$$

for sufficiently large L and Y_{k_i} ($k_i\in\mathbb{N}, i=1,\ldots,4$) as defined in (A4). Furthermore, Lemma 3.4 remains valid with $g=h_c$ since the function h_c is continuous, which is a consequence of the inequality

$$\begin{aligned}
&|h_c(x,y) - h_c(\bar{x},\bar{y})| \\
&\leq f(x,\bar{x},y,\bar{y})\,[\|x-\bar{x}\|_1+\|y-\bar{y}\|_1] + \int_{\mathbb{R}^d}\left|\widetilde{h}^{(c)}(x,y)-\widetilde{h}^{(c)}(\bar{x},y)\right|P_X(dy) \\
&\quad + \int_{\mathbb{R}^d}\left|\widetilde{h}^{(c)}(x,y)-\widetilde{h}^{(c)}(x,\bar{y})\right|P_X(dy) \\
&\leq f(x,\bar{x},y,\bar{y})\,[\|x-\bar{x}\|_1+\|y-\bar{y}\|_1] + C\,P(\|X_1\|_1>z) \\
&\quad + \int_{\mathbb{R}^d} f(x,\bar{x},y,y)\,\mathbb{1}_{\|y\|_1\leq z}\,P_X(dy)\,\|x-\bar{x}\|_1 + \int_{\mathbb{R}^d} f(x,x,y,\bar{y})\,\mathbb{1}_{\|x\|_1\leq z}\,P_X(dx)\,\|y-\bar{y}\|_1.
\end{aligned} \tag{3.19}$$

Therefore, one can follow the lines of step 1 of the proof of Lemma 3.5 and plug in inequality (3.18) above. This procedure leads to $\limsup_{n\to\infty} n^2\,\mathbb{E}(U_{n,c}-U_{n,c}^{(L)})^2 \longrightarrow_{L\to\infty} 0$

whenever $(J(L))_L$ is chosen appropriately.
Step 3: $\lim_{K\to\infty} \limsup_{n\to\infty} n^2\, \mathbb{E}(U_{n,c}^{(L)} - U_{n,c}^{(K,L)})^2 = 0$.
First, Lipschitz continuity of the function $h_c^{(K,L)}$ implies

$$\begin{aligned}
&\mathbb{E}\left|H^{(K)}(Y_{k_1}, Y_{k_2}) - H^{(K)}(Y_{k_3}, Y_{k_4})\right| \\
&\leq \left(\mathbb{E}|f_4(Y_{k_1}, Y_{k_2}, Y_{k_3}, Y_{k_4})|^{1/(1-\delta)} \sum_{i=1}^{4} \|Y_i\|_1\right)^{1-\delta} (\mathbb{E}\|Y_{k_1} - Y_{k_3}\|_1 + \mathbb{E}\|Y_{k_2} - Y_{k_4}\|_1)^{\delta}
\end{aligned}$$

with $f_4 := C + f_3$ and $H^{(K)} = h_c^{(L)} - h_c^{(K,L)}$. Under the assumption (A4)(i), the inequality

$$\mathbb{E}\left|H^{(K)}(Y_{k_1}, Y_{k_2}) - H^{(K)}(Y_{k_3}, Y_{k_4})\right| \leq C[\mathbb{E}(\|Y_{k_1} - Y_{k_3}\|_1 + \|Y_{k_2} - Y_{k_4}\|_1)]^{\delta}$$

holds true if $J(L)$ is chosen sufficiently large, cf. step 2 of this proof. Imitating step 2 of the proof of Lemma 3.5, one obtains $\limsup_{n\to\infty} n^2\, \mathbb{E}(U_{n,c}^{(L)} - U_{n,c}^{(K,L)})^2 \longrightarrow_{K\to\infty} 0$. Summing up the three steps leads to

$$\lim_{c\to\infty} \lim_{L\to\infty} \lim_{K\to\infty} \limsup_{n\to\infty} n^2\, \mathbb{E}\left(U_n - U_{n,c}^{(K,L)}\right)^2 = 0.$$

□

This auxiliary result implies the analogue to Theorem 3.1 for non-Lipschitz kernels.

Theorem 3.2. *Assume that the conditions (A1), (A2), and (A4) are satisfied. Then, as $n \to \infty$,*

$$n\, U_n \xrightarrow{d} Z,$$

where Z is defined as in Theorem 3.1. If additionally $\mathbb{E}|h(X_1, X_1)| < \infty$, then, as $n \to \infty$,

$$n\, V_n \xrightarrow{d} Z + \mathbb{E}h(X_1, X_1).$$

Proof. In complete accordance with the proof of Theorem 3.1 we get $n\, U_n \xrightarrow{d} Z$ on the basis of Lemma 3.8. In order to obtain $n\, V_n \xrightarrow{d} Z + \mathbb{E}h(X_1, X_1)$, convergence of $n^{-1} \sum_{k=1}^{n} h(X_k, X_k)$ to $\mathbb{E}h(X_1, X_1)$, in probability, has to be verified. This is a consequence of Lemma 2.3 since h is Lipschitz continuous on any bounded interval. □

4 Bootstrap consistency for degenerate U- and V-statistics

4.1 Motivation

The bootstrap is a powerful method to estimate the distributions of various statistics. However, its validity usually has to be verified in a case-by-case manner. Asymptotic validity or consistency means that the distribution of the statistic of interest and the distribution of its bootstrap counterpart converge in a certain metric. Besides the aforementioned case-by-case considerations, the literature also provides some results that apply to general frameworks. Examples include the work of Bickel and Freedman [12] on Efron's bootstrap for different statistics of i.i.d. observations, of Stute, Gonzáles Manteiga and Presedo Quindimil [105] for bootstrapping the empirical process with estimated parameters for i.i.d. data, and of Chen and Romano [23] for tests in the frequency domain concerning the validity of time series models. In this chapter we establish a quite general result on bootstrap consistency for statistics of degenerate U- and V-type when the underlying random variables are weakly dependent.

As it is demonstrated in the preceding part of the thesis, the asymptotic distributions of degenerate U- and V-statistics of independent as well as dependent data are basically that of infinite weighted sums of products of normal random variables. These statistics often emerge from goodness-of-fit tests. When testing composite hypotheses, the weights of the limiting variable depend on unknown parameters. Thus, the tests are in general not asymptotically distribution-free and (asymptotic) critical values cannot be obtained directly. The bootstrap offers a capable alternative to estimate these quantities. Several results have been derived in the context of i.i.d. observations. A naïve application of Efron's bootstrap fails since the summands of U- and V-statistics are not independent even though the underlying sample variables are. Furthermore, the kernel of the bootstrap statistics is no longer degenerate. Arcones and Giné [5] circumvented these difficulties by degenerating the bootstrap counterpart of the function h artificially. That is,

$$\begin{aligned} h_{BOOT}(x,y) := h(x,y) - \int h(x,y) P_n(dy) - \int h(x,y) P_n(dx) \\ + \iint h(x,y) P_n(dx)\, P_n(dy), \end{aligned} \tag{4.1}$$

where P_n denotes the empirical distribution of the sample under consideration. They verified bootstrap consistency for the statistics associated with the function h_{BOOT} by a special decomposition of the kernel and applying the bootstrap central limit theorem of Bickel and Freedman [12]. Dehling and Mikosch [40] derived the same result via proving convergence in a modified Mallows' metric. Their approach is based on coupling techniques. Both methods basically rely on moment constraints concerning the function h. Jiménez-Gamero, Muñoz-García and Pino-Mejías [77] considered statistics with estimated parameters and allowed for parametric bootstrap methods. They restricted themselves to kernels of the form $h(x, y; \theta) = \int q(x, t; \theta)\, q(y, t; \theta)\, G(dt)$ that often result from test statistics of L_2-type. Here, G denotes a finite measure over $\mathbb{R}^d$. Besides certain moment constraints, they imposed regularity conditions on the density of the underlying sample. Imitating the proof for the asymptotic behaviour of the original statistics, i.e. employing a spectral decomposition of h, they obtained distributional convergence of the bootstrap statistics to the same limit. Due to continuity of the distribution function of the limiting variable, consistency w.r.t. the uniform metric can be deduced. Leucht and Neumann [86] considered statistics with parametrized kernels, too. Applying a coupling technique similar to the work of Dehling and Mikosch [40], the authors obtained consistency of general bootstrap methods. Their approach does not require smoothness assumptions on the density that are difficult to check in cases where the density has no closed form. Instead, additional moment constraints on the kernel were imposed on the bootstrap side.

To the author's best knowledge, there are no results available regarding bootstrap for general degenerate U- and V-type statistics of dependent data so far. In the subsequent section we will fill this gap. Coupling techniques as they were applied by Dehling and Mikosch [40] or Leucht and Neumann [86] cannot be invoked here since they heavily rely on the independence of the underlying sample variables. Rather we re-derive the asymptotic distributions of the statistics on the bootstrap side. Additionally, a sufficient condition for the continuity of the limiting distribution function is established, which in turn yields bootstrap consistency.

4.2 Deriving bootstrap consistency

In order to obtain bootstrap consistency when the underlying observations are dependent, one has to assure that the bootstrap procedure captures the dependence structure. Therefore, we first introduce a bootstrap counterpart of the τ-dependence coefficient. Given $X_1, \ldots, X_n$, let X^* and Y^* denote vectors of bootstrap random variables with values in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, respectively. To describe the dependence structure of the bootstrap sample, we define $\mathbb{X}_n := (X_1', \ldots, X_n')'$ and, in analogy to Definition 2.2,

$$\tau^*(Y^*, X^*, x_n) :=$$
$$\mathbb{E}\left(\sup_{f \in \Lambda_1(\mathbb{R}^{d_1}, \|\cdot - \cdot\|_1)} \left| \int_{\mathbb{R}^{d_1}} f(x) P_{X^*|Y^*}(dx) - \int_{\mathbb{R}^{d_1}} f(x) P_{X^*}(dx) \right| \,\Big|\, \mathbb{X}_n = x_n \right),$$

provided that $\mathbb{E}(\|X^*\|_1 \mid \mathbb{X}_n = x_n) < \infty$. We make the following assumption:

(**A1***) **(i)** The sequence of bootstrap variables is stationary with probability tending to one. Additionally, $(X_{t_1}^{*\prime}, X_{t_2}^{*\prime})' \xrightarrow{d} (X_{t_1}', X_{t_2}')',\ \forall\, t_1, t_2 \in \mathbb{N}$, holds true in probability.

(ii) Conditionally on $X_1, \ldots, X_n$, the random variables $(X_n^*)_{n\in\mathbb{N}}$ are τ-weakly dependent, i.e. there exist some constant $C_1 < \infty$, a sequence of coefficients $(\bar{\tau}_r)_{r\in\mathbb{N}}$ with $\sum_{r=1}^{\infty} r(\bar{\tau}_r)^\delta < \infty$ for some $\delta \in (0,1)$, and a sequence of sets $(\mathfrak{X}_n^{(1)})_{n\in\mathbb{N}}$ with $P(\mathbb{X}_n \in \mathfrak{X}_n^{(1)}) \longrightarrow_{n\to\infty} 1$ and the following property: For any sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in \mathfrak{X}_n^{(1)},\ n \in \mathbb{N}$, the inequality $\sup_{k\in\mathbb{N}} \mathbb{E}(\|X_k^*\|_1 \mid \mathbb{X}_n = x_n) \le C_1$ holds true. Moreover, the coefficients

$$\tau_r^*(x_n) := \sup\big\{\tau^*\big((X_{s_1}^{*\prime}, \ldots, X_{s_u}^{*\prime})', (X_{t_1}^{*\prime}, X_{t_2}^{*\prime}, X_{t_3}^{*\prime})', x_n\big) \mid \\ u \in \mathbb{N},\ s_1 \le \ldots \le s_u < s_u + r \le t_1 \le t_2 \le t_3\ \in \mathbb{N}\big\}$$

can be bounded by $\bar{\tau}_r$ for all $r \in \mathbb{N}$.

These conditions are fairly natural. Loosely speaking, we merely require that the bootstrap algorithm mimics the distribution and the dependence structure of the original sample. Sufficient conditions for the convergence of the fidis are provided in Subsection 4.3.1. Note that we do not assume the bootstrap variables to be stationary for large n. This constraint would be violated in cases when the parameter estimators only fall with probability tending to one into the set of parameters that assures stationarity of the corresponding process. Typical examples are the model-based AR(p) and ARCH(p) bootstrap methods considered by Neumann and Paparoditis [94].

Under the assumption above, we already obtain distributional convergence of the bootstrap counterparts of $n\,U_n$ and $n\,V_n$ towards the limiting variable $Z(+\mathbb{E}h(X_1, X_1))$ for a confined class of kernels.

Lemma 4.1. *Suppose that (A1) and (A1*) hold true. Further let $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a bounded, symmetric, Lipschitz continuous function such that*
$\mathbb{E}h(X_1, y) = \mathbb{E}(h(X_1^*, y)|X_1, \ldots, X_n) = 0,\ \forall\, y \in \mathbb{R}^d$. *Then, as $n \to \infty$,*

$$\frac{1}{n-1}\sum_{\substack{j,k=1\\ j\neq k}}^{n} h(X_j^*, X_k^*) \xrightarrow{d} Z \quad \textit{and} \quad \frac{1}{n}\sum_{j,k=1}^{n} h(X_j^*, X_k^*) \xrightarrow{d} Z + \mathbb{E}h(X_1, X_1)$$

hold in probability. Here, Z is defined as in Theorem 3.1.

Proof. This assertion is an immediate consequence of Theorem 4.1 below. □

In order to deduce bootstrap consistency, convergence in a certain metric ρ is additionally required, i.e.

$$\rho\Big(P\Big(\frac{1}{n-1}\sum_{\substack{j,k=1\\ j\neq k}}^{n} h(X_j^*, X_k^*) \le x \big| X_1, ..., X_n\Big),\ P\Big(\frac{1}{n-1}\sum_{\substack{j,k=1\\ j\neq k}}^{n} h(X_j, X_k) \le x\Big)\Big) \xrightarrow{P} 0.$$

Convergence in the uniform metric follows from Lemma 4.1 if the limit distribution has a continuous cumulative distribution function. The next result states a necessary and sufficient condition for this.

Lemma 4.2. *The distribution of the limit variable Z, derived in Theorem 3.1 / Theorem 3.2 under (A1), (A2), and (A3)/(A4), has a continuous cumulative distribution function if* $\operatorname{var}(Z) > 0$.

Proof. According to Lemma 3.1, Lemma 3.5 and Lemma 3.8, respectively,

$$\lim_{c\to\infty} \lim_{L\to\infty} \lim_{K\to\infty} \limsup_{n\to\infty} n^2\, \mathbb{E}\left(U_n - U_{n,c}^{(K,L)}\right)^2 = 0$$

which we verified to imply

$$\lim_{c\to\infty} \lim_{L\to\infty} \lim_{K\to\infty} \mathbb{E}\left(Z - Z_c^{(K,L)}\right)^2 = 0,$$

cf. the proof of Theorem 3.1. Thus, a positive variance of Z yields the existence of constants $V > 0$ and $c_0 > 0$ such that for each $c \geq c_0$ we can find some $L_0 \in \mathbb{N}$ such that for every $L \geq L_0$ there is a $K_0 \in \mathbb{N}$ with $\operatorname{var}(Z_c^{(K,L)}) \geq V$, $\forall\, K \geq K_0$. Moreover, uniform equicontinuity of the distribution functions of the corresponding family $(((Z_c^{(K,L)})_K)_L)_c$ yields the desired property of Z since for all $\varepsilon > 0$,

$$P(Z \leq x_0) - P(Z < x_0) \leq \limsup_{c\to\infty} \limsup_{L\to\infty} \limsup_{K\to\infty} [P(Z_c^{(K,L)} \leq x_0 + \varepsilon) - P(Z_c^{(K,L)} \leq x_0 - \varepsilon)].$$

By matrices-based notation we obtain the following representation:

$$\begin{aligned} Z_c^{(K,L)} &= \sum_{k_1,k_2 \in \{-K,\dots,K\}^d} \alpha_{k_1,k_2}^{(c)} [Z_{k_1} Z_{k_2} - A_{k_1,k_2}] \\ &\quad + \sum_{j=0}^{J(L)-1} \sum_{k_1,k_2 \in \{-K,\dots,K\}^d} \sum_{e=(e_1',e_2')' \in \bar{E}} \beta_{j;k_1,k_2}^{(c,e)} [Z_{j;k_1}^{(e_1)} Z_{j;k_2}^{(e_2)} - B_{j;k_1,k_2}^{(e)}] \\ &= \sum_{k,l=1}^{M(K,L)} \gamma_{k,l}^{(c,K,L)} \bar{Z}_k \bar{Z}_l + C_{k,l} = Z^{(K,L)\prime}\, \Gamma_c^{(K,L)}\, Z^{(K,L)} + C^{(K,L)} \end{aligned}$$

with a certain symmetric matrix of coefficients $\Gamma_c^{(K,L)} = [\gamma_{k,l}^{(c,K,L)}]_{k,l=1}^{M(K,L)}$, constants $C_{k,l}$ and $C^{(K,L)}$, as well as a vector $Z^{(K,L)} = (\bar{Z}_1, \dots, \bar{Z}_{M(K,L)})' \sim N\left(0_{M(K,L)}, \Sigma^{(K,L)}\right)$ with a specific matrix $\Sigma^{(K,L)}$. The positive semidefinite, symmetric covariance matrix can be expressed as $\Sigma^{(K,L)} = O^{(K,L)} S^{(K,L)} [O^{(K,L)}]'$, where $O^{(K,L)}$ denotes a certain orthogonal matrix and $S^{(K,L)}$ a specified diagonal matrix. Hence, $Z_c^{(K,L)}$ can be rewritten as

$$Z_c^{(K,L)} \stackrel{d}{=} \bar{Y}' \left(\left[S^{(K,L)}\right]^{1/2} O^{(K,L)\prime}\, \Gamma_c^{(K,L)}\, O^{(K,L)} \left[S^{(K,L)}\right]^{1/2}\right) \bar{Y} + C^{(K,L)}$$

with a standard normal random vector $\bar{Y}$. Note that the matrix in round brackets is symmetric and thus diagonalizable. Consequently, there exist a certain orthogonal matrix $U_c^{(K,L)}$ and a matrix $\Lambda_c^{(K,L)} := diag(\lambda_1^{(c,K,L)}, ..., \lambda_{M(K,L)}^{(c,K,L)})$ with $|\lambda_1^{(c,K,L)}| \geq ... \geq |\lambda_{M(K,L)}^{(c,K,L)}|$ such that

$$\begin{aligned} Z_c^{(K,L)} &\overset{d}{=} \bar{Y}' \, U_c^{(K,L)\prime} \Lambda_c^{(K,L)} U_c^{(K,L)} \bar{Y} + C^{(K,L)} \\ &= Y' \, \Lambda_c^{(K,L)} Y + C^{(K,L)} \\ &= \sum_{k=1}^{M(K,L)} \lambda_k^{(c,K,L)} Y_k^2 + C^{(K,L)}. \end{aligned}$$

Here, $Y = \big(Y_1, \ldots, Y_{M(K,L)}\big)'$ is a multivariate standard normally distributed random vector. For notational simplicity we suppress the upper index (c, K, L) in the sequel. Due to the above choice of the triple (c, K, L), either $\sum_{k=1}^{4} \lambda_k^2$ or $\sum_{k=5}^{M(K,L)} \lambda_k^2$ is bounded from below by $V/4$.

In the first case, $|\lambda_1| \geq \sqrt{V}/4$ holds true which implies

$$P\big(Z_c^{(K,L)} \in [x - \varepsilon, x + \varepsilon]\big) \leq \int_0^{2\varepsilon} f_{|\lambda_1| Y_1^2}(t) \, dt \leq P\big(Y_1^2 \leq 2\varepsilon\big) \, \max\left\{1, \frac{4}{\sqrt{V}}\right\}$$

for arbitrary $\varepsilon > 0, x \in \mathbb{R}$, where $f_{|\lambda_1| Y_1^2}$ denotes the probability density of $|\lambda_1| Y_1^2$. Here, the first inequality results from the fact that convolution preserves the continuity properties of the smoother function and from the monotonicity of the density of a χ^2 distributed random variable.

In the opposite case, i.e. $\sum_{k=5}^{M(K,L)} \lambda_k^2 \geq V/4$, it is possible to bound the uniform norm of the density function of $Z_c^{(K,L)}$ by means of its variance. To this end, we first consider the characteristic function $\varphi_{Z_c^{(K,L)}}$ of $Z_c^{(K,L)}$ and assume w.l.o.g. that $M(K, L)$ is divisible by four. Defining a sequence $(\mu_k)_{k=1}^{M(K,L)/4}$ by $\mu_k = \lambda_{4k}$ for $k \in \{1, ..., M(K, L)/4\}$ allows for the following estimation:

$$\begin{aligned} \left|\varphi_{Z_c^{(K,L)}}(t)\right| &= \prod_{j=1}^{M(K,L)} \left|(1 \,-\, 2i\lambda_j t)^{-1/2}\right| \\ &= \left\{ \prod_{j=1}^{M(K,L)} \big(1 \,+\, (2\lambda_j t)^2\big) \right\}^{-1/4} \\ &\leq \left\{ \prod_{j=1}^{M(K,L)/4} \big(1 \,+\, (2\mu_j t)^2\big) \right\}^{-1} \\ &\leq \frac{1}{1 \,+\, 4\,(\mu_1^2 + \cdots + \mu_{M(K,L)/4}^2)\, t^2}. \end{aligned}$$

By inverse Fourier transform we obtain the subsequent result concerning the density function of $Z_c^{(K,L)}$:

$$
\begin{aligned}
\left\| f_{Z_c^{(K,L)}} \right\|_\infty &\le \frac{1}{2\pi} \int_{\mathbb{R}} |\varphi_{Z_c^{(K,L)}}(t)|\, dt \\
&\le \frac{1}{2\pi} \int_{\mathbb{R}} \frac{1}{1 + (2\sqrt{\mu_1^2 + \cdots + \mu_{M(K,L)/4}^2}\, t)^2}\, dt \\
&= \frac{1}{\sqrt{\mu_1^2 + \cdots + \mu_{M(K,L)/4}^2}} \frac{1}{2\pi} \int_0^\infty \frac{1}{1+u^2}\, du \\
&\le \frac{1}{2\sqrt{4(\mu_1^2 + \cdots + \mu_{M(K,L)/4-1}^2)}} \\
&\le \frac{1}{2\sqrt{\lambda_5^2 + \cdots + \lambda_{M(K,L)}^2}} \\
&\le \frac{1}{\sqrt{V}}.
\end{aligned}
$$

Thus, $P(Z_c^{(K,L)} \in [x-\varepsilon, x+\varepsilon]) \le 2\varepsilon/\sqrt{V}$, which completes the studies of the case $\sum_{k=5}^{M(K,L)} \lambda_k^2 > V/4$ and finally yields the assertion. □

Remark 4.1. In most applications the condition of a positive variance of the limit distribution, that assures continuity of the asymptotic distribution function, has to be verified in a case-by-case manner. Nevertheless, it turns out to be a standard assumption in similar contexts, see for example Shao and Yu [101], Theorem 2.4, or Dehling and Wendler [42], Theorem 2.1. Also in the paper by Chen and Romano [23], Theorem 3.3, continuity of the limiting distribution function was not accessible in general.

Kernels of statistics emerging from goodness-of-fit tests for composite hypotheses often depend on an unknown parameter, cf. Chapter 5. We intend to employ bootstrap consistency for this setting, i.e. when parameters have to be estimated. Moreover, the kernels resulting from L_2-test statistics are in general neither Lipschitz continuous nor bounded. In the next chapter, we establish L_2-tests based on the empirical characteristic function, which is bounded and Lipschitz continuous. However, when approximating the test statistic by a statistic of degenerate V-type, a linearization of the parameter estimator is invoked. In general, the function of linearization is neither bounded nor Lipschitz continuous. Thus, even for these simple kinds of tests, Lemma 4.1 does not apply. We need to enlarge the class of feasible kernels. For this purpose we additionally assume:

(A2*) **(i)** The sequence of parameter estimators satisfies $\widehat{\theta}_n \xrightarrow{P} \theta \in \Theta \subseteq \mathbb{R}^p$.

(ii) The kernel is degenerate w.r.t. $P^*_{X_1^*}$[1], i.e. $\mathbb{E}^* h(X_1^*, y, \widehat{\theta}_n) = 0,\ \forall\, y \in \mathbb{R}^d$.

(iii) For some δ satisfying (A1*)(ii), $\nu > (2-\delta)/(1-\delta)$, and a constant $C_2 < \infty$, there exists a sequence of sets $(\mathfrak{X}_n^{(2)})_{n\in\mathbb{N}}$ such that $P(\mathbb{X}_n \in \mathfrak{X}_n^{(2)}) \longrightarrow_{n\to\infty} 1$ and

[1] Throughout the thesis, the expressions P^* and $\mathbb{E}^*$ denote the bootstrap distribution and bootstrap expectation conditionally on $X_1, \ldots, X_n$.

for all $(x_n)_{n\in\mathbb{N}}$ with $x_n \in \mathfrak{X}_n^{(2)}$, $n \in \mathbb{N}$, the moment constraint

$$\sup_{1\leq k<n} \mathbb{E}\big(|h(X_1^*, X_{1+k}^*, \widehat{\theta}_n)|^\nu + |h(X_1^*, \widetilde{X}_1^*, \widehat{\theta}_n)|^\nu | \mathbb{X}_n = x_n\big) \leq C_2$$

holds true, where (conditionally on $X_1, \ldots, X_n$) $\widetilde{X}_1^*$ denotes an independent copy of X_1^*.

and concerning the regularity of the kernel:

(**A3***) **(i)** The kernel is continuous in its third argument in a neighbourhood $U(\theta) \subseteq \Theta$ of θ and satisfies

$$\left|h(x, y, \widehat{\theta}_n) - h(\bar{x}, \bar{y}, \widehat{\theta}_n)\right| \leq f(x, \bar{x}, y, \bar{y}, \widehat{\theta}_n)\, [\|x - \bar{x}\|_1 + \|y - \bar{y}\|_1]$$

for all x, $\bar{x}$, y, $\bar{y} \in \mathbb{R}^d$, where the function $f : \mathbb{R}^{4d} \times \mathbb{R}^p \to \mathbb{R}$ is invariant under permutations of its first four arguments and continuous on $\mathbb{R}^{4d} \times U(\theta)$. Moreover, for $\eta := 1/(1-\delta)$ with δ of (A2*), some $A > 0$, and some $C_3 < \infty$, there exists a sequence of sets $(\mathfrak{X}_n^{(3)})_{n\in\mathbb{N}}$ such that $P(\mathbb{X}_n \in \mathfrak{X}_n^{(3)}) \longrightarrow_{n\to\infty} 1$ and for all $(x_n)_{n\in\mathbb{N}}$ with $x_n \in \mathfrak{X}_n^{(3)}$, $n \in \mathbb{N}$, the moment constraint

$$\sup_{1\leq k_1,\ldots,k_5\leq n} \mathbb{E}\Big\{ \max_{a_1,a_2\in[-A,A]^d} \big[f(Y_{k_1}^*, Y_{k_2}^* + a_1, Y_{k_3}^*, Y_{k_4}^* + a_2, \widehat{\theta}_n)\big]^\eta \\ \times \|Y_{k_5}^*\|_1 | \mathbb{X}_n = x_n\Big\} \leq C_3$$

holds true for any vector $(Y_{k_1}^{*\prime}, \ldots, Y_{k_5}^{*\prime})'$ consisting of independent subvectors $(Y_{k_{j_1(m)}}^{*\prime}, \ldots, Y_{k_{j_l(m)}}^{*\prime})' \stackrel{d}{=} (X_{k_{j_1(m)}}^{*\prime}, \ldots, X_{k_{j_l(m)}}^{*\prime})'$, $l, m = 1, \ldots, 5$, conditionally on $X_1, \ldots, X_n$.

(ii) The dependence coefficients satisfy $\sum_{r=1}^\infty r(\bar{\tau}_r)^{\delta^2} < \infty$.

Remark 4.2. (i) The assumptions above basically assure that smoothness and moment constraints of h carry over from the original variables to the bootstrap side. In this sense, the assumptions are comparable to those made by Leucht and Neumann [86] in the i.i.d. case.

(ii) Modifying the bootstrap statistics allows for omitting the condition (A2*)(ii), cf. Subsection 4.3.2.

(iii) If $h(\cdot, \cdot, \bar{\theta})$ is Lipschitz continuous uniformly for all $\bar{\theta}$ in a neighbourhood of θ, then (A3*) can be omitted.

Now the bootstrap statistics

$$n\, U_n^* := \frac{1}{n-1} \sum_{\substack{j,k=1\\ j\neq k}}^n h(X_j^*, X_k^*, \widehat{\theta}_n) \quad \text{and} \quad n\, V_n^* := \frac{1}{n} \sum_{j,k=1}^n h(X_j^*, X_k^*, \widehat{\theta}_n)$$

are investigated. To this end, we denote the U- and V-statistics with kernel $h(x,y) = h_\theta(x,y) = h(x,y,\theta)$ and arguments $X_1,\dots,X_n$ by U_n and V_n, respectively. The limiting random variable Z of Theorem 3.2 depends on the unknown parameter θ through the coefficients

$$\alpha_{k_1,k_2}^{(c)} = \alpha_{k_1,k_2}^{(c)}(\theta) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c(x,y,\theta)\,\Phi_{0,k_1}(x)\,\Phi_{0,k_2}(y)\,dx\,dy$$

and

$$\beta_{j;k_1,k_2}^{(c,e)} = \beta_{j;k_1,k_2}^{(c,e)}(\theta) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c(x,y,\theta)\,\Psi_{j;k_1,k_2}^{(e)}(x,y)\,dx\,dy,\ j\in\mathbb{Z},\ k_1,k_2\in\mathbb{Z}^d,\ e\in\bar{E},$$

where h_c is defined as in Section 3.2. Under the assumptions above we establish bootstrap consistency.

Theorem 4.1. *Suppose that the conditions (A1), (A2), and (A4) as well as (A1*), (A2*), and (A3*) are fulfilled.*

(i) As $n\to\infty$,

$$n\,U_n^* \xrightarrow{d} Z, \quad \textit{in probability,}$$

where Z is defined as in Theorem 3.1. If furthermore $\operatorname{var}(Z)>0$, *then*

$$\sup_{-\infty<x<\infty} |P(n\,U_n^*\le x|X_1,\dots,X_n) - P(n\,U_n\le x)| \xrightarrow{P} 0.$$

(ii) If additionally $\mathbb{E}|h(X_1,X_1,\theta)|<\infty$ and $\mathbb{E}(|h(X_1^*,X_1^*,\widehat{\theta}_n)|\mid\mathbb{X}_n) \xrightarrow{P} \mathbb{E}|h(X_1,X_1,\theta)|$, then, as $n\to\infty$,

$$n\,V_n^* \xrightarrow{d} Z + \mathbb{E}h(X_1,X_1,\theta), \quad \textit{in probability.}$$

Moreover, in case of $\operatorname{var}(Z)>0$,

$$\sup_{-\infty<x<\infty} |P(n\,V_n^*\le x|X_1,\dots,X_n) - P(n\,V_n\le x)| \xrightarrow{P} 0.$$

Remark 4.3. This theorem implies that bootstrap-based tests of U- or V-type have asymptotically a prescribed size α, i.e. $P(n\,U_n > t_{u,\alpha}^*) \longrightarrow_{n\to\infty} \alpha$ and $P(n\,V_n > t_{v,\alpha}^*) \longrightarrow_{n\to\infty} \alpha$, where $t_{u,\alpha}^*$ and $t_{v,\alpha}^*$ denote the $(1-\alpha)$-quantiles of $n\,U_n^*$ and $n\,V_n^*$, respectively, given $X_1,\dots,X_n$.

Proof. *Step 1: Approximation of U-statistics.*
By virtue of Lemma 4.2 it suffices to verify distributional convergence. Convergence in the uniform norm then follows by common arguments from monotonicity and boundedness of the distribution function. Based on the definition of the events $\mathfrak{X}_n^{(1)}$, $\mathfrak{X}_n^{(2)}$, and $\mathfrak{X}_n^{(3)}$ within the assumptions (A1*), (A2*), and (A3*), respectively, we introduce

$$\mathfrak{X}_n^\theta \subseteq \mathfrak{X}_n^{(1)}\cap\mathfrak{X}_n^{(2)}\cap\mathfrak{X}_n^{(3)}\cap\left\{\mathbb{X}_n \,\middle|\, \|\widehat{\theta}_n-\theta\|_1<\delta_n\right\}$$

such that

$$P_{X^*_{t_1},\dots,X^*_{t_k}|\mathbb{X}_n=x_n} = P_{X^*_{t_1+l},\dots,X^*_{t_k+l}|\mathbb{X}_n=x_n}, \tag{4.2}$$

$$P_{X^*_{t_1},X^*_{t_2}|\mathbb{X}_n=x_n} \Longrightarrow P_{X_{t_1},X_{t_2}} \tag{4.3}$$

uniformly for any sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in \mathfrak{X}^\theta_n$, where $t_1,\dots,t_k,\ k,l\in\mathbb{N}$. Moreover, the null sequence $(\delta_n)_{n\in\mathbb{N}}$ can be chosen such that $\widehat{\theta}_n(x_n)\in U(\theta)$ for all $x_n\in\mathfrak{X}^\theta_n,\ n\in\mathbb{N}$, and $P(\mathbb{X}_n\in\mathfrak{X}^\theta_n)\longrightarrow_{n\to\infty} 1$ holds. Hence, to prove $n\,U^*_n \xrightarrow{d} Z$, in probability, it suffices to verify convergence in distribution of $n\,U^*_n$ to Z conditionally on $\mathbb{X}_n = x_n$ for any sequence $(x_n)_n$ with $x_n\in\mathfrak{X}^\theta_n,\ n\in\mathbb{N}$. Now, we take an arbitrary sequence $(x_n)_n$ with $x_n\in\mathfrak{X}^\theta_n,\ n\in\mathbb{N}$.

In order to show that it suffices to analyse statistics with bounded kernels, we consider the w.r.t. $P_{X^*_1|\mathbb{X}_n=x_n}$ degenerate version h^*_c of

$$\widetilde{h}^*_c(x,y,\widehat{\theta}_n) := \begin{cases} h(x,y,\widehat{\theta}_n) & \text{for } |h(x,y,\widehat{\theta}_n)|\le c_h(\widehat{\theta}_n),\\ -c_h(\widehat{\theta}_n) & \text{for } h(x,y,\widehat{\theta}_n) < -c_h(\widehat{\theta}_n),\\ c_h(\widehat{\theta}_n) & \text{for } h(x,y,\widehat{\theta}_n) > c_h(\widehat{\theta}_n),\end{cases}$$

where $c_h(\widehat{\theta}_n) := \max_{x,y\in[-c,c]^d}|h(x,y,\widehat{\theta}_n)|$ is uniformly bounded on $(\mathfrak{X}^\theta_n)_{n\in\mathbb{N}}$. The associated U-statistic is referred to as $U^*_{n,c}$. Proceeding as in the proof of Lemma 3.1 and as in step 1 of the proof of Lemma 3.8 results in

$$\limsup_{n\to\infty} n^2\,\mathbb{E}\left[\left(U^*_n - U^*_{n,c}\right)^2|\mathbb{X}_n = x_n\right]\le \varepsilon_c$$

with $\varepsilon_c \longrightarrow_{c\to\infty} 0$. Within the calculations the relation $\limsup_{n\to\infty} P\left(X^*_1\notin(-c,c)^d \mid \mathbb{X}_n = x_n\right) \le P\left(X_1\notin(-c,c)^d\right) \longrightarrow_{c\to\infty} 0$ has to be invoked, which follows from (4.3) in conjunction with Portmanteau's theorem.

Next, we approximate the bounded kernel by the degenerate version $\widehat{h}^{*(K,L)}_c$ of

$$\begin{aligned}\widetilde{h}^{*(K,L)}_c := &\sum_{k_1,k_2\in\{-K,\dots,K\}^d} \widehat{\alpha}^{(c)}_{k_1,k_2}\Phi_{0,k_1}\Phi_{0,k_2}\\ &+\sum_{j=0}^{J(L)-1}\ \sum_{k_1,k_2\in\{-K,\dots,K\}^d}\ \sum_{e\in\bar{E}}\widehat{\beta}^{(c,e)}_{j;k_1,k_2}\Psi^{(e)}_{j;k_1,k_2},\end{aligned}$$

where

$$\widehat{\alpha}^{(c)}_{k_1,k_2} := \iint_{\mathbb{R}^d\times\mathbb{R}^d} h^*_c(x,y,\widehat{\theta}_n)\,\Phi_{0,k_1}(x)\,\Phi_{0,k_2}(y)\,dx\,dy,$$

$$\widehat{\beta}^{(c,e)}_{j;k_1,k_2} := \iint_{\mathbb{R}^d\times\mathbb{R}^d} h^*_c(x,y,\widehat{\theta}_n)\,\Psi^{(e)}_{j;k_1,k_2}(x,y)\,dx\,dy,$$

and the associated U-statistic is denoted by $\widehat{U}^{*(K,L)}_{n,c}$. For this purpose first note that $c_h(\widehat{\theta}_n)\longrightarrow_{n\to\infty} c_h(\theta)$, which denotes the corresponding truncation parameter in the definition of $\widetilde{h}^{(c)}(x,y,\theta)$. This implies in conjunction with (4.3) and $\widehat{\theta}_n \longrightarrow_{n\to\infty}\theta$ that $h^*_c(x,y,\widehat{\theta}_n)\longrightarrow_{n\to\infty} h_c(x,y,\theta)$ for all $x,y\in\mathbb{R}^d$. Invoking inequality (3.19) of the proof of Lemma 3.8 and its bootstrap counterpart, we additionally get uniform convergence on any compact interval. Moreover, the relation

$$\widehat{\alpha}^{(c)}_{J;k_1,k_2}(x_n) := \iint_{\mathbb{R}^d\times\mathbb{R}^d} h^*_c(x,y,\widehat{\theta}_n(x_n))\,\Phi_{J,k_1}(x)\,\Phi_{J,k_2}(y)\,dx\,dy \longrightarrow_{n\to\infty} \alpha^{(c)}_{J;k_1,k_2}$$

leads to uniform convergence on any compact interval of $\widetilde{h}^{*(L)} := \sum_{k_1,k_2\in\mathbb{Z}^d} \widehat{\alpha}^{(c)}_{J(L);k_1,k_2}\,\Phi_{J(L),k_1}\,\Phi_{J(L),k_2}$ towards $\widetilde{h}^{(L)}_c$, defined in the proof of Lemma 3.5. Based on these preliminary considerations, we are in the position to prove

$$\lim_{L\to\infty}\lim_{K\to\infty}\limsup_{n\to\infty} n^2\,\mathbb{E}\left[\left(U^*_{n,c} - \widehat{U}^{*(K,L)}_{n,c}\right)^2 \,\middle|\, \mathbb{X}_n = x_n\right] = 0$$

by following the lines of the proof of Lemma 3.8. Here, $J(L)$ is chosen as follows: Since $\limsup_{n\to\infty} P(X^*_1 \notin (-b,b)^d|\mathbb{X}_n = x_n) \le P(X_1 \notin (-b,b)^d)$, we first select some $b = b(L) < \infty$ such that $P(X_1 \notin (-b,b)^d) \le 1/L$. Afterwards, one determines $J(L)$ such that $\max_{x,y\in[-b,b]^d} |h_c(x,y,\theta) - \widetilde{h}^{(L)}_c(x,y,\theta)| \le 1/L$ and $S_\phi/2^{J(L)} < A$, where S_ϕ denotes the length of the support of the scale function ϕ and the constant A is defined as in assumption (A3*). In view of our foregoing considerations, this in turn implies $\limsup_{n\to\infty}\max_{x,y\in[-b,b]^d} |h^*_c(x,y,\widehat{\theta}_n) - \widetilde{h}^{*(L)}_c(x,y,\widehat{\theta}_n)| \le 1/L$ and thus the desired result of convergence.

Based on the relations $\widehat{\alpha}^{(c)}_{k_1,k_2}(x_n) \longrightarrow_{n\to\infty} \alpha^{(c)}_{k_1,k_2}$ and $\widehat{\beta}^{(c,e)}_{j;k_1,k_2}(x_n) \longrightarrow_{n\to\infty} \beta^{(c,e)}_{j;k_1,k_2}$, one can prove

$$n^2\,\mathbb{E}\left[\left(\widehat{U}^{*(K,L)}_{n,c} - U^{*(K,L)}_{n,c}\right)^2 \,\middle|\, \mathbb{X}_n = x_n\right] \underset{n\to\infty}{\longrightarrow} 0,$$

where the kernel $h^{*(K,L)}_c$ of $U^{*(K,L)}_{n,c}$ is obtained by substituting $\widehat{\alpha}^{(c)}_{k_1,k_2}$ and $\widehat{\beta}^{(c,e)}_{j;k_1,k_2}$ in the kernel of $\widehat{U}^{*(K,L)}_{n,c}$ through $\alpha^{(c)}_{k_1,k_2}$ and $\beta^{(c,e)}_{j;k_1,k_2}$. This result can be justified invoking the decomposition of $\mathbb{E}[n\,Z_n]^2$ considered in (3.2) with $H := H_n = \widehat{h}^{*(K,L)}_c - h^{*(K,L)}_c$. Of course, the approach has to be slightly modified since we are not given a sequence of random variables but a triangular scheme here. In fact, the conditional expectation $\mathbb{E}([n\,Z_n]^2 \mid \mathbb{X}_n = x_n)$ has to be bounded in the present context. The resulting sum is split up in the same manner as in (3.2). Exemplarily, we briefly investigate the triangular version of $\limsup_{n\to\infty}(n-1)^{-2}\sum_{r=1}^{n-1} Z^{(1)}_{n,r}$ and $\sup_{k\in\mathbb{N}}\mathbb{E}(H^2_n(X^*_1, X^*_{1+k}) \mid \mathbb{X}_n = x_n)$. Convergence of the latter term immediately follows from the convergence of the coefficients. The summands of the modified version of $Z^{(1)}_{n,r}$ are given by

$$\left|\mathbb{E}(H_n(X^*_i,X^*_j)H_n(X^*_k,X^*_l) \mid \mathbb{X}_n = x_n) - \mathbb{E}(H_n(X^*_i,\widetilde{X}^*_j)H_n(\widetilde{X}^*_k,\widetilde{X}^*_l) \mid \mathbb{X}_n = x_n)\right|. \quad (4.4)$$

Here, $(\widetilde{X}^{*\prime}_j, \widetilde{X}^{*\prime}_k, \widetilde{X}^{*\prime}_l)'$ is a copy of $(X^{*\prime}_j, X^{*\prime}_k, X^{*\prime}_l)'$ that is independent of X^*_i (conditionally on $\mathbb{X}_n = x_n$) and chosen such that $\mathbb{E}(\|(\widetilde{X}^{*\prime}_j, \widetilde{X}^{*\prime}_k, \widetilde{X}^{*\prime}_l) - (X^{*\prime}_j, X^{*\prime}_k, X^{*\prime}_l)\|_1 \mid \mathbb{X}_n = x_n) \le \bar{\tau}_r$. (This may possibly require an enlargement of the underlying probability space, cf. Lemma 2.1.) Now, the expression (4.4) can be approximated from above by

$$C(\bar{\tau}_r)^\delta\left[\mathbb{E}\left(|H_n(X^*_k,X^*_l)|^{1/(1-\delta)} \mid \mathbb{X}_n = x_n\right) + \mathbb{E}\left(|H_n(X^*_i,\widetilde{X}^*_j)|^{1/(1-\delta)} \mid \mathbb{X}_n = x_n\right)\right]^{1-\delta}$$

due to uniform boundedness and uniform Lipschitz continuity of the sequence of functions $(H_n)_n$. Thus we obtain that the triangular version of $(n-1)^{-2}\sum_{r=1}^{n-1} Z_{n,r}^{(1)}$ can be bounded from above by $o(1)\sum_{r=1}^{n-1}(\bar{\tau}_r)^\delta$ which vanishes asymptotically.

Step 2: *Asymptotics for U-statistics.*

We rewrite the approximating statistic $n\,U_{n,c}^{*(K,L)}$ as follows:

$$\begin{aligned} &n\,U_{n,c}^{*(K,L)}\\ &= \frac{1}{n-1}\sum_{i\neq j} h_c^{*(K,L)}(X_i^*, X_j^*)\\ &= \frac{n}{n-1}\sum_{k,l=1}^{M(K,L)} \gamma_{k,l}^{(c)} \left(\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n q_k^*(X_i^*)\right]\left[\frac{1}{\sqrt{n}}\sum_{j=1}^n q_l^*(X_j^*)\right] - \frac{1}{n}\sum_{i=1}^n q_k^*(X_i^*) q_l(X_i^*)\right), \end{aligned}$$

where q_k^* denotes the centered version of $\widetilde{q}_k$ (w.r.t. $P_{X_1^*|\mathbb{X}_n=x_n}$) and the sequence $(\widetilde{q}_k)_k$ is defined as in the proof of Lemma 3.6. The latter term in the round brackets converges to $\mathbb{E}q_k(X_1)q_l(X_1)$ on $\mathfrak{X}_n^\theta$ by virtue of Lemma 2.3 because of the Lipschitz continuity and boundedness of the functions q_k^* and the validity of $\sup_{n\in\mathbb{N}} P(\|X_1^*\|_1 > K \mid \mathbb{X}_n = x_n) \longrightarrow_{K\to\infty} 0$.

For the analysis of the remaining quantities, we introduce $Q_i^* := \sum_{k=1}^{M(K,L)} t_k\, q_k^*(X_i^*)$, $t_1, ..., t_{M(K,L)} \in \mathbb{R}$. Obviously, given $\mathbb{X}_n = x_n$, the row-wise stationary triangular scheme $(Q_i^*)_i$ consists of centered and uniformly bounded random variables that satisfy the dependence condition of Lemma 2.2. To apply this result, it remains to show that the inequality $|n^{-1}\operatorname{var}(Q_1^* + \cdots + Q_n^*|\mathbb{X}_n = x_n) - \sigma^2| < \varepsilon, \forall\, n \geq n_0(\varepsilon)$, holds true for arbitrary $\varepsilon > 0$ with $\sigma^2 = \operatorname{var}(Q_1) + 2\sum_{r=2}^\infty \operatorname{cov}(Q_1, Q_r)$ and $(Q_k)_k$ as in the proof of Lemma 3.6. To this end, the abbreviations $\operatorname{var}^*(\cdot) = \operatorname{var}(\cdot|\mathbb{X}_n = x_n)$ and $\operatorname{cov}^*(\cdot) = \operatorname{cov}(\cdot|\mathbb{X}_n = x_n)$ are used. We have

$$\begin{aligned} &\left|\frac{1}{n}\operatorname{var}^*[Q_1^* + \cdots + Q_n^*] - \sigma^2\right|\\ &\leq 2\sum_{r=2}^\infty \min\left\{\frac{r-1}{n}, 1\right\} |\operatorname{cov}^*(Q_1^*, Q_r^*)| + \left|\operatorname{var}^*(Q_1^*) + 2\sum_{r=2}^\infty \operatorname{cov}^*(Q_1^*, Q_r^*) - \sigma^2\right|\\ &\leq 2\sum_{r=2}^\infty \min\left\{\frac{r-1}{n}, 1\right\} |\operatorname{cov}^*(Q_1^*, Q_r^*)| + 2\left|\sum_{r=2}^{R-1}[\operatorname{cov}^*(Q_1^*, Q_r^*) - \operatorname{cov}(Q_1, Q_r)]\right|\\ &\quad + |\operatorname{var}^*(Q_1^*) - \operatorname{var}(Q_1)| + 2\left|\sum_{r\geq R}\operatorname{cov}^*(Q_1^*, Q_r^*)\right| + 2\left|\sum_{r\geq R}\operatorname{cov}(Q_1, Q_r)\right|. \end{aligned}$$

By (A1), (A1*), and boundedness of Q_1 as well as Q_1^*, the index R can be chosen such that $|\sum_{r\geq R}\operatorname{cov}(Q_1, Q_r)| + |\sum_{r\geq R}\operatorname{cov}^*(Q_1^*, Q_r^*)| \leq \varepsilon/4$. Moreover, the condition (A1*) implies that the first summand can be bounded from above by $\varepsilon/4$ as well if $n \geq n_0(\varepsilon)$ for some sufficiently large $n_0(\varepsilon) \in \mathbb{N}$. According to the convergence of the two-dimensional distributions and the uniform boundedness of $(Q_k^*)_{k\in\mathbb{N}}$, it is possible to pick $n_0(\varepsilon)$ such that additionally the two remaining summands are bounded by $\varepsilon/4$. Eventually, we checked all prerequisites of Lemma 2.2 and thus obtain the assertion of the central limit theorem.

The application of the continuous mapping theorem results in $n\,U_{n,c}^{*(K,L)} \xrightarrow{d} Z_c^{(K,L)}$, in probability. This in turn implies $n\,U_n^* \xrightarrow{d} Z$, in probability, by the same arguments as in the proof of Theorem 3.1.

Step 3: Asymptotics for V-statistics.
Define $\widetilde{\mathfrak{X}}_n^\theta \subseteq \mathfrak{X}_n^\theta$, $n \in \mathbb{N}$, such that

$$|\mathbb{E}(|h(X_1^*, X_1^*, \widehat{\theta}_n)| \mid \mathbb{X}_n = x_n) - \mathbb{E}|h(X_1, X_1, \theta)|| \leq \eta_n, \quad \forall\, x_n \in \widetilde{\mathfrak{X}}_n^\theta.$$

Here, the null sequence $(\eta_n)_{n\in\mathbb{N}}$ can be chosen in such a way that $P(\mathbb{X}_n \in \widetilde{\mathfrak{X}}_n^\theta) \longrightarrow_{n\to\infty} 1$. In order to obtain the desired result of convergence for $n\,V_n^*$, additionally to our previous investigations,

$$P\left(\left|\frac{1}{n}\sum_{k=1}^n h(X_k^*, X_k^*, \widehat{\theta}_n) - \mathbb{E}h(X_1, X_1, \theta)\right| > \varepsilon \,\middle|\, \mathbb{X}_n = x_n\right) \underset{n\to\infty}{\longrightarrow} 0$$

has to be proved for arbitrary $\varepsilon > 0$ and any sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in \widetilde{\mathfrak{X}}_n^\theta$, $n \in \mathbb{N}$. The definition of the sets $(\widetilde{\mathfrak{X}}_n^\theta)_n$ implies uniform integrability of $h(X_1^*, X_1^*, \widehat{\theta}_n)$ w.r.t. $P_{X_1^*|X_n=x_n}$. Thus, we get $\mathbb{E}\big(h(X_1^*, X_1^*, \widehat{\theta}_n) \mid \mathbb{X}_n = x_n\big) \longrightarrow_{n\to\infty} \mathbb{E}h(X_1, X_1, \theta)$ and it remains to verify

$$P\left(\left|\frac{1}{n}\sum_{k=1}^n \Big[h(X_k^*, X_k^*, \widehat{\theta}_n) - \mathbb{E}(h(X_1^*, X_1^*, \widehat{\theta}_n)|\mathbb{X}_n = x_n)\Big]\right| > \frac{\varepsilon}{2} \,\middle|\, \mathbb{X}_n = x_n\right) \underset{\mathrm{n}\to\infty}{\longrightarrow} 0,$$

This in turn is a consequence of Lemma 2.3 since under the assumptions of the theorem, $\sup_{n\in\mathbb{N}} P(\|X_1^*\|_1 > K \mid \mathbb{X}_n = x_n) \longrightarrow_{K\to\infty} 0$ and the sequence of functions $(g_n)_{n\in\mathbb{N}}$ with $g^{(n)}(\cdot) = h(\cdot, \cdot, \widehat{\theta}_n(x_n)) - \mathbb{E}(h(X_1^*, X_1^*, \widehat{\theta}_n)|\mathbb{X}_n = x_n)$ is uniformly integrable and Lipschitz continuous on any bounded interval. Finally, bootstrap consistency follows from Lemma 4.2. □

4.3 Some notes on the assumptions

4.3.1 Convergence of the finite-dimensional distributions

A consistent bootstrap procedure has to capture the covariance structure of the underlying sample. To guarantee this property, we assume convergence of the two-dimensional distributions of the bootstrap sample towards those of the original sample. Of course, this is a very weak assumption if the observations are independent and identically distributed. In particular, Efron's bootstrap satisfies this condition. In general, it is not self-evident how to check assumption (A1*)(i) within applications. Neumann and Paparoditis [94] were faced with a similar problem when deriving a bootstrap-aided test for Markovian time series models. They found out that in this context the key for convergence of the fidis is convergence of the conditional distributions. Here, the metric

$$d(P,Q) = \inf_{X\sim P,\, Y\sim Q} \mathbb{E}\left(\|X - Y\| \wedge 1\right)$$

between two distributions on $(\mathbb{R}^m, \mathcal{B}^m)$ has been used, where $\|\cdot\|$ is any norm on $\mathbb{R}^m$.

Lemma 4.3 (Neumann and Paparoditis [94]). *Assume that* $(Y_t)_{t\in\mathbb{Z}}$ *and* $(Y_t^{(n)})_{t\in\mathbb{Z}}$, $n \in \mathbb{N}$, *are stationary Markov processes of order* p, *defined on probability spaces* $(\Omega, \mathcal{A}, P)$ *and* $(\Omega^{(n)}, \mathcal{A}^{(n)}, P^{(n)})$, *respectively. Further suppose that*

(i) for all compact sets $K \subseteq \mathbb{R}^p$,

$$\sup_{y\in K} d\left(P^{(n)}_{Y_t^{(n)}|Y_{t-1}^{(n)}=y_1,\ldots,Y_{t-p}^{(n)}=y_p}, P_{Y_t|Y_{t-1}=y_1,\ldots,Y_{t-p}=y_p}\right) \underset{n\to\infty}{\longrightarrow} 0,$$

(ii) for all $y \in \mathbb{R}^p$,

$$\sup_{\widetilde{y}:\|\widetilde{y}-y\|\leq\delta} d\left(P_{Y_t|Y_{t-1}=y_1,\ldots,Y_{t-p}=y_p}, P_{Y_t|Y_{t-1}=\widetilde{y}_1,\ldots,Y_{t-p}=\widetilde{y}_p}\right) \underset{\delta\to 0}{\longrightarrow} 0,$$

(iii) $(P^{(n)}_{Y_t^{(n)}})_{n\in\mathbb{N}}$ *is tight and*

(iv) there is a unique stationary distribution $P_{Y_1,\ldots,Y_p}$ *that corresponds to* $P_{Y_t|Y_{t-1},\ldots,Y_{t-p}}$.

Then, for all $k \in \mathbb{N}$, $P^{(n)}_{Y_1^{(n)},\ldots,Y_k^{(n)}} \Longrightarrow P_{Y_1,\ldots,Y_k}$.

In particular, they showed that AR(p) bootstrap and ARCH(p) bootstrap yield samples that satisfy (A1*)(i), cf. Section 4.4 of this thesis.

Remark 4.4. Neumann and Paparoditis [94] stated the result for univariate Markov processes. However, the proof given in the corresponding preprint, Neumann and Paparoditis [93], remains valid if the underlying processes have values in $\mathbb{R}^d$.

Bühlmann [21] obtained convergence of the fidis in the case of a sieve bootstrap procedure for linear processes, i.e. $X_t - \mathbb{E}(X_1) = \sum_{k=0}^{\infty} a_k\,\varepsilon_{t-k}$ with $a_0 = 1$ and a sequence $(\varepsilon_k)_k$ of i.i.d. centered random variables. His approach is based on an approximation of the linear process by AR(p) processes with increasing order $p = p(n) \longrightarrow_{n\to\infty} \infty$ such that $\widehat{\varepsilon}_{n,t} = \sum_{j=0}^{p(n)} \widehat{\phi}_{n,j}(X_{t-j} - \bar{X})$, $t = p+1,\ldots,n$. Here, $\widehat{\phi}_{n,0} = 1$, $(\widehat{\phi}_{n,k})_{k=1}^p$ are the corresponding parameter estimates, and $\bar{X}$ denotes the sample mean. One proceeds with ordinary autoregressive bootstrap. That means, in a first step the bootstrap innovations $(\varepsilon_t^*)_t$ are drawn with replacement from the sample of the re-centered versions of $(\widehat{\varepsilon}_{n,t})_{t=p+1}^n$. Then the bootstrap sample is defined recursively by $\varepsilon_t^* = \sum_{j=0}^{p(n)} \widehat{\phi}_{n,j}(X^*_{t-j} - \bar{X})$.

Lemma 4.4 (Bühlmann [21]). *Let* $(X_k)_k$ *be a linear process as above with* $\mathbb{E}|\varepsilon_1|^4 < \infty$. *Suppose that the coefficients* $(a_k)_k$ *are absolutely summable and that* $\sum_{k=0}^{\infty} a_k z^k$, $z \in \mathbb{C}$, *has no root within the unit circle. Moreover, let* $p = p(n) = o\left([n/(log(n)]^{1/4}\right)$ *and assume that* $\widehat{\phi}_p = (\widehat{\phi}_{n,1},\ldots,\widehat{\phi}_{n,p})'$ *satisfy the empirical Yule-Walker equations* $\widehat{\Gamma}_p\,\widehat{\phi}_p = -\widehat{\gamma}_p$, *where* $\widehat{\Gamma}_p = [\widehat{R}(i-j)]_{i,j=1}^p$, $\widehat{\gamma}_p = (\widehat{R}(1),\ldots,\widehat{R}(p))'$ *and* $\widehat{R}(j) = n^{-1}\sum_{t=1}^{n-|j|}(X_t - \bar{X})(X_{t+|j|} - \bar{X})$. *Then, for every* $t_1,\ldots,t_k \in \mathbb{N}$, $k \in \mathbb{N}$,

$$P_{Y^*_{t_1},\ldots,Y^*_{t_k}} \Longrightarrow P_{Y_{t_1},\ldots,Y_{t_k}}, \quad \textit{in probability.}$$

4.3.2 Degeneracy of the bootstrap kernel

Under the assumption (A2)(i), the condition (A2*)(ii) postulates that the function h, that is degenerate w.r.t. P_X, additionally meets the degeneracy condition w.r.t. the bootstrap distribution $P_{X_1^*|\mathbb{X}_n}$. Using model-based bootstrap methods in the i.i.d. setting, this condition is usually satisfied, see Leucht and Neumann [86] for a typical example. However, it turns out that the application of Efron's bootstrap contradicts the assumption of degeneracy. In this case, substituting the kernel h by the function h_{BOOT}, defined in equation (4.1) of Section 4.1, in the bootstrap statistics induces consistency.

Both situations, i.e. the condition (A2*)(ii) is or is not fulfilled under validity of the assumption (A2)(i), occur when the underlying observations are weakly dependent. The bootstrap counterparts of U- and V-statistics considered in the framework of a model-specification test in Section 5.5 are degenerate. In Section 5.4 we extend the characteristic function-based goodness-of-fit test considered by Leucht and Neumann [86] for i.i.d. data towards dependent random variables. While in the independent case degeneracy of the bootstrap statistic can be easily obtained, the bootstrap counterpart of the corresponding test statistic in the dependent setting is not degenerate in most cases. Motivated by the above-named approach of Arcones and Giné [5] and Dehling and Mikosch [40], we degenerate the bootstrap test statistic artificially. Actually, the assumption (A2*)(ii) can be omitted if, instead of $n\,U_n^*$ and $n\,V_n^*$, we apply the statistics $n\,\bar{U}_n^*$ and $n\,\bar{V}_n^*$ characterized by the artificially degenerated kernel

$$\begin{aligned}&\bar{h}(X_i^*, X_j^*, \widehat{\theta}_n)\\ &:= h(X_i^*, X_j^*, \widehat{\theta}_n) - \int_{\mathbb{R}^d} h(x, X_j^*, \widehat{\theta}_n)\, P_{X_1^*|\mathbb{X}_n}(dx)\\ &\quad - \int_{\mathbb{R}^d} h(X_i^*, y, \widehat{\theta}_n)\, P_{X_1^*|\mathbb{X}_n}(dy) + \iint_{\mathbb{R}^d\times\mathbb{R}^d} h(x, y, \widehat{\theta}_n)\, P_{X_1^*|\mathbb{X}_n}(dx)\, P_{X_1^*|\mathbb{X}_n}(dy) \quad \text{a.s.,}\end{aligned}$$

$i, j = 1, \ldots, n$, which is well-defined with probability tending to one due to (A2*)(iii).

Proposition 4.1. *Suppose that the assumptions (A1), (A2), and (A4) as well as (A1*), (A2*)(i),(iii), and (A3*) are satisfied.*

(i) As $n \to \infty$,

$$n\,\bar{U}_n^* \xrightarrow{d} Z, \quad \textit{in probability,}$$

where Z is defined as in Theorem 3.1. If furthermore $\operatorname{var}(Z) > 0$, *then*

$$\sup_{-\infty<x<\infty} \big|P(n\,\bar{U}_n^* \le x | X_1, ..., X_n) - P(n\,U_n \le x)\big| \xrightarrow{P} 0.$$

(ii) If additionally $\mathbb{E}(|h(X_1^*, X_1^*, \widehat{\theta}_n)| \mid \mathbb{X}_n) \xrightarrow{P} \mathbb{E}|h(X_1, X_1, \theta)| < \infty$, then, as $n \to \infty$,

$$n\,\bar{V}_n^* \xrightarrow{d} Z + \mathbb{E}h(X_1, X_1, \theta), \quad \textit{in probability.}$$

Moreover, in case of $\operatorname{var}(Z) > 0$,

$$\sup_{-\infty<x<\infty} \big|P(n\,\bar{V}_n^* \le x | X_1, ..., X_n) - P(n\,V_n \le x)\big| \xrightarrow{P} 0.$$

Remark 4.5. Although the (conditional) bootstrap distribution $P_{X_1^*|\mathbb{X}_n}$ is commonly unknown, it can always be approximated arbitrarily well by simulation. That implies that the bootstrap statistics can indeed be simulated.

In order to verify the assertion, only a few modifications of the proof of Theorem 4.1 are necessary. This results from the fact that in the step of reducing the problem to statistics with bounded kernels in the former proof, we already worked with artificially degenerated functions.

Proof. Let $\mathfrak{X}_n^\theta$ be defined as in the proof of Theorem 4.1 and $(x_n)_{n\in\mathbb{N}}$ an arbitrary sequence with $x_n \in \mathfrak{X}_n^\theta,\ n \in \mathbb{N}$. We first analyse the moment structure and continuity properties of $\bar{h}$, which is well-defined on $\mathfrak{X}_n^\theta$. According to Minkowski's and Jensen's inequalities and the definition of $\mathfrak{X}_n^{(2)} \supseteq \mathfrak{X}_n^\theta$ in assumption (A2*)(iii), the function $\bar{h}$ exhibits all moments considered in (A2*)(iii). Additionally, we have

$$\begin{aligned}
&|\bar{h}(Y_{k_1}^*, Y_{k_2}^*, \widehat{\theta}_n) - \bar{h}(Y_{k_3}^*, Y_{k_4}^*, \widehat{\theta}_n)| \\
&\leq \Big\{ f(Y_{k_1}^*, Y_{k_3}^*, Y_{k_2}^*, Y_{k_4}^*, \widehat{\theta}_n) + \int_{\mathbb{R}^d} f(Y_{k_1}^*, Y_{k_3}^*, z, z, \widehat{\theta}_n)\, P_{X_1^*|\mathbb{X}_n = x_n}(dz) \\
&\quad + \int_{\mathbb{R}^d} f(z, z, Y_{k_2}^*, Y_{k_4}^*, \widehat{\theta}_n)\, P_{X_1^*|\mathbb{X}_n = x_n}(dz) \Big\} \left[\|Y_{k_1}^* - Y_{k_3}^*\|_1 + \|Y_{k_2}^* - \bar{Y}_{k_4}^*\|_1 \right] \\
&=: \bar{f}_n(Y_{k_1}^*, Y_{k_3}^*, Y_{k_2}^*, Y_{k_4}^*, \widehat{\theta}_n) \left[\|Y_{k_1}^* - Y_{k_3}^*\|_1 + \|Y_{k_2}^* - Y_{k_4}^*\|_1 \right]
\end{aligned}$$

for any vector $(Y_{k_1}^{*\prime}, Y_{k_2}^{*\prime}, Y_{k_3}^{*\prime}, Y_{k_4}^{*\prime})'$ defined in (A3*). Invoking Minkowski's inequality and Jensen's inequality, we get that $\bar{f}_n$ inherits all moments of f stated in (A3*)(i).

Based on these considerations, we are in the position to reduce the problem to statistics with bounded kernels. For this purpose we apply the truncated kernels $\widetilde{h}_c^*$, their degenerate versions h_c^*, and the associated U-statistics $U_{n,c}^*,\ c \in \mathbb{R}_+$, that are equivalently defined as in the proof of Theorem 4.1. To establish an upper bound for $n^2\, \mathbb{E}[(\bar{U}_n^* - U_{n,c}^*)^2 | \mathbb{X}_n = x_n]$, we proceed as in the proofs of Lemma 3.1 and Lemma 3.8, respectively. The only differences are that now we plug in the bootstrap variables into the calculations, use $\bar{H}_c = \bar{h} - h_c^*$, and substitute the integrals

$$\int_{\mathbb{R}^d} \widetilde{h}^{(c)}(X_k, y)\, P_X(dy), \quad \int_{\mathbb{R}^d} \widetilde{h}^{(c)}(y, X_l)\, P_X(dy) \text{ and } \iint_{\mathbb{R}^d\times\mathbb{R}^d} \widetilde{h}^{(c)}(x, y)\, P_X(dx)\, P_X(dy)$$

by

$$\begin{aligned}
&\int_{\mathbb{R}^d} \widetilde{h}_c^*(X_k^*, y, \widehat{\theta}_n) - h(X_k^*, y, \widehat{\theta}_n)\, P_{X_1^*|\mathbb{X}_n = x_n}(dy), \\
&\int_{\mathbb{R}^d} \widetilde{h}_c^*(x, X_l^*, \widehat{\theta}_n) - h(y, X_l^*, \widehat{\theta}_n)\, P_{X_1^*|\mathbb{X}_n = x_n}(dy)
\end{aligned}$$

and

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \widetilde{h}_c^*(x, y, \widehat{\theta}_n) - h(x, y, \widehat{\theta}_n)\, P_{X_1^*|\mathbb{X}_n = x_n}(dx)\, P_{X_1^*|\mathbb{X}_n = x_n}(dy)$$

in the analysis of E_1 defined in the proof of Lemma 3.1. This procedure finally implies

$$\lim_{c\to\infty} \limsup_{n\to\infty} n^2\, \mathbb{E}\left[(\bar{U}_n^* - U_{n,c}^*)^2 | \mathbb{X}_n = x_n \right] = 0$$

for all sequences $(x_n)_n$ with $x_n \in \mathfrak{X}_n^\theta$. Obviously, all remaining steps to verify $n\,\bar{U}_n^* \xrightarrow{d} Z$, in probability, are identical to those of the proof of Theorem 4.1.

In order to deduce the convergence of the corresponding V-type statistics, an arbitrary sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in \widetilde{\mathfrak{X}}_n^\theta$, $n \in \mathbb{N}$, is considered, where $(\widetilde{\mathfrak{X}}_n^\theta)_{n\in\mathbb{N}}$ is defined as in step 3 of the proof of Theorem 4.1. According to the investigations there, the definition of $\bar{h}$, and the relation $\mathbb{E}(h(X_1^*, \widetilde{X}_1^*, \widehat{\theta}_n) \mid \mathbb{X}_n = x_n) \longrightarrow_{n\to\infty} \mathbb{E}h(X_1, \widetilde{X}_1, \theta) = 0$, we obtain

$$
\begin{aligned}
&P\Big(\Big|\frac{1}{n}\sum_{k=1}^n \Big[\bar{h}(X_k^*, X_k^*, \widehat{\theta}_n) - \mathbb{E}(h(X_1, X_1, \theta))\Big]\Big| > \varepsilon \Big| \mathbb{X}_n = x_n\Big) \\
&\leq P\Big(\Big|\frac{2}{n}\sum_{k=1}^n \Big[\int_{\mathbb{R}^d} h(X_k^*, y, \widehat{\theta}_n)\, P_{X_1^*|\mathbb{X}_n = x_n}(dy) \\
&\qquad\qquad - \mathbb{E}(h(X_1^*, \widetilde{X}_1^*, \widehat{\theta}_n)|\mathbb{X}_n = x_n)\Big]\Big| > \frac{\varepsilon}{2}\Big|\mathbb{X}_n = x_n\Big) + o(1).
\end{aligned}
$$

Here, $\widetilde{X}_1^*$ and $\widetilde{X}_1$ denote independent copies of X_1^* and X_1, respectively. Lemma 2.3 can not be employed directly to verify that the remaining term tends to zero as n increases since the involved functions satisfy all its prerequisites except the uniform Lipschitz continuity on bounded intervals. However, this condition is only required once in the proof of our version of the weak law of large numbers, namely to bound $\mathbb{E}\left|g_{M,K}^{(n,c)}(X_{n,k}) - g_{M,K}^{(n,c)}(\widetilde{X}_{n,k})\right|$ by $C(M,K)\,\bar{\tau}_{k-j}$, cf. inequality (2.10). Here, $\widetilde{X}_{n,k}$ denotes a copy of $X_{n,k}$ that is independent of $X_{n,j}$ for a specified $j < k$. Adopting the notation with $g^{(n)}(X_{n,k}) = \int_{\mathbb{R}^d} h(X_k^*, y, \widehat{\theta}_n) P_{X_1^*|\mathbb{X}_n = x_n}(dy) - \mathbb{E}(h(X_1^*, \widetilde{X}_1^*, \widehat{\theta}_n)|\mathbb{X}_n = x_n)$ and $(X_{n,k})_{k=1}^n$ with $(X_{n,1}', \ldots, \mathbb{X}_{n,n}')' \sim P_{(X_1^{*\prime}, \ldots, \mathbb{X}_n^{*\prime})'|\mathbb{X}_n = x_n}$, we obtain

$$
\begin{aligned}
&\mathbb{E}\left|g_{M,K}^{(n,c)}(X_{n,k}) - g_{M,K}^{(n,c)}(\widetilde{X}_{n,k})\right| \\
&\leq \mathbb{E}\Big(\Big[\left|g^{(n)}(X_{n,k})\right| + \left|g^{(n)}(\widetilde{X}_{n,k})\right|\Big]\, \|w_K(X_{n,k}) - w_K(\widetilde{X}_{n,k})\|_1 \\
&\quad + \int_{\mathbb{R}^d} \left|f(X_{n,k}, \widetilde{X}_{n,k}, y, y, \widehat{\theta}_n)\right| P_{X_1^*|\mathbb{X}_n = x_n}(dy)\, \|X_{n,k} - \widetilde{X}_{n,k}\|_1\Big) \\
&\leq C(K)\left[1 + \mathbb{E}\left(|h(X_1^*, \widetilde{X}_1^*, \widehat{\theta}_n)|^{1-\delta} \mid \mathbb{X}_n = \omega_n\right) + C_3\right]^{1/(1-\delta)} \bar{\tau}_r^\delta
\end{aligned}
$$

in the present case by Hölder's inequality, (A2*), and (A3*). This implies that the assertion of Lemma 2.3 remains valid, which finally completes the proof. □

4.4 Examples

We verified bootstrap consistency under quite general assumptions regarding the bootstrap counterpart of the underlying sample. Below certain procedures are listed that satisfy the condition (A1*). We have to check three criteria, namely (1) stationarity, (2) convergence of the fidis, and (3) τ-dependence with the associated summability condition on the coefficients.

4.4.1 AR(p) Bootstrap

The autoregressive bootstrap certainly belongs to the most known model-based bootstrap methods. Employing results of Neumann and Paparoditis [94], we establish the validity of (A1*) for this procedure.

Let $X = (X_t)_{t\in\mathbb{Z}}$ be an AR(p) process, i.e.

$$X_t = \sum_{i=1}^{p} \theta_i \, X_{t-i} + \varepsilon_t,$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. centered innovations with finite variance. The vector $(\theta_1, \ldots, \theta_p)'$ is assumed to be located in the parameter space

$$\Theta := \left\{ \theta \in \mathbb{R}^p \,\middle|\, 1 - \sum_{i=1}^{p} \theta_i \, z^i \neq 0 \;\; \forall\, z \in \mathbb{C} \text{ with } |z| \leq 1 \right\}, \tag{4.5}$$

which assures the existence of a unique stationary solution to the associated AR(p) model equation. According to Example 2.1.1 of Subsection 2.1.2 this process is τ-weakly dependent with exponentially decaying coefficients.

Following Neumann and Paparoditis [94], we consider the subsequent parametric bootstrap procedure based on the observations $X_{1-p}, \ldots, X_n$.

[B1] Algorithm.

1. Calculate a consistent estimator $\widehat{\theta}_n = (\widehat{\theta}_{n,1}, \ldots, \widehat{\theta}_{n,p})'$ of θ.
2. Compute the estimated residuals $\widetilde{\varepsilon}_t = X_t - \sum_{i=1}^{p} \widehat{\theta}_{n,i} \, X_{t-i}$ and their centered versions $\widehat{\varepsilon}_t = \widetilde{\varepsilon}_t - n^{-1} \sum_{k=1}^{n} \widetilde{\varepsilon}_k$.
3. Draw bootstrap innovations via Efron's bootstrap from the empirical distribution function $\widehat{F}_\varepsilon(\cdot) = n^{-1} \sum_{k=1}^{n} \mathbb{1}_{\widehat{\varepsilon}_k \leq \cdot}$.
4. Determine an initial vector $(X_0^*, \ldots, X_{1-p}^*)'$.
5. Calculate the bootstrap variables $X_t^* = \sum_{i=1}^{p} \widehat{\theta}_{n,i} \, X_{t-i}^* + \varepsilon_t^*, \; t = 1, 2, \ldots$.

Lemma 4.5. *Let $(X_t)_t$ be an AR(p) process with $\theta \in \Theta$, where Θ is defined as in (4.5). A bootstrap sample generated by the Algorithm [B1] meets the condition (A1*) if the probability density of the innovations is bounded and if the initial vector is drawn from the stationary bootstrap distribution, provided its existence.*

Proof. (1) Within the proof of their Lemma 4.3, Neumann and Paparoditis [94] verified that the roots of the polynomial $\widehat{\theta}(z) = 1 - \widehat{\theta}_{n,1} z - \cdots - \widehat{\theta}_{n,p} z^p$ lie outside the unit circle with probability tending to one. This yields the existence of a unique stationary solution to $Y_t = \sum_{i=1}^{p} \widehat{\theta}_{n,i} Y_{t-i} + \varepsilon_t^*, \; t \in \mathbb{Z}$. If we draw the initial vector of the bootstrap process from this distribution, we obtain a stationary process.

(2) Let the bootstrap process be started with its stationary distribution (provided its existence). Convergence of the fidis has been proved by Neumann and Paparoditis [94] employing Lemma 4.3 under the assumption of the innovations to exhibit a bounded probability density.

(3) According to the proof of Lemma 4.3 of Neumann and Paparoditis [94], the bootstrap process has, with probability tending to one, the MA(∞) representation $X_t^* = \sum_{k=0}^\infty \widehat{a}_{n,k}\, \varepsilon_{t-k}^*$ with $|\widehat{a}_{n,k}| \leq C(1-\delta)^{-k}$ for some $\delta \in (0,1)$ and $\mathbb{E}((\varepsilon_1^*)^2 | \mathbb{X}_n) \leq K < \infty$. Thus, our Example 2.1.3 yields the τ-dependence property with the corresponding summability condition.

$\square$

Remark 4.6. To prove the above lemma, we assumed the starting value to be drawn from the stationary bootstrap distribution, which exists with probability tending to one. In general, this distribution is unknown. However, the influence of the starting value on X_n^* decreases rapidly with increasing n if the roots of the characteristic polynomial lie outside the unit ball. Thus, we expect the theory to hold true in the general framework, too. In practice one would start the algorithm with a "reasonable" initial vector and drop the first generated values. Similarly we can proceed to tackle this minor difficulty in all subsequent examples.

4.4.2 ARCH(p) Bootstrap

The class of autoregressive conditional heteroscedastic processes was introduced by Engle [58] in order to describe the temporal development of the volatility of the inflation in the United Kingdom. A model-based bootstrap method for this type of process that satisfies the assumption (A1*) can be found in the paper of Neumann and Paparoditis [94] again.

Let $X = (X_t)_{t \in Z}$ be a stationary process that satisfies the ARCH model equation

$$X_t = \sqrt{\theta_0 + \theta_1 X_{t-1}^2 + \cdots + \theta_p X_{t-p}^2}\, \varepsilon_t, \tag{4.6}$$

where $\theta \in \Theta$ and

$$\Theta = \left\{ \theta = (\theta_0, \theta_1, \ldots, \theta_p)' \in \mathbb{R}^{p+1} \mid \theta_0 > 0,\ \theta_i \geq 0,\ i = 1, \ldots, p, \text{ and } \sum_{i=1}^p \theta_i < 1 \right\}. \tag{4.7}$$

Moreover, $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a sequence of centered i.i.d. innovations with unit variance. Based on standard normal innovations, Milhøj [89] derived a unique stationary solution $\widetilde{X}_t = \varepsilon_t [\theta_0 \sum_{k=0}^\infty M(t,k)]^{1/2}$ of (4.6). Here, $M(t,0) = 1$ and $M(t,k) = \sum_{a_1,\ldots,a_k=1}^p \prod_{i=1}^k \theta_{a_i} \varepsilon_{t-a_1-\cdots-a_k}^2$ with $\mathbb{E}M(t,k) = (\sum_{i=1}^p \theta_i)^k$. His investigations can be carried over directly to innovations with the above characteristics. Additionally, note that by virtue of Example 2.1.2 the stationary process is τ-weakly dependent with exponentially decreasing coefficients since $\mathbb{E}X_1^2 < \infty$ by $\mathbb{E}M(t,k) = (\sum_{i=1}^p \theta_i)^k$.

Analogously to Neumann and Paparoditis [94], we consider the following model-based bootstrap procedure conditionally on the observations $X_{1-p}, \ldots, X_n$.

[B2] Algorithm.

1. Calculate a consistent estimator $\widehat{\theta}_n = (\widehat{\theta}_{n,0}, \ldots, \widehat{\theta}_{n,p})'$ of θ.
2. Compute the estimated residuals $\widetilde{\varepsilon}_t = X_t / \sqrt{\theta_0 + \theta_1 X_{t-1}^2 + \cdots + \theta_p X_{t-p}^2}$ and their standardized versions $\widehat{\varepsilon}_t = \bar{\varepsilon}_t / \sqrt{n^{-1} \sum_{k=1}^n \bar{\varepsilon}_k^2}$, where $\bar{\varepsilon}_t = \widetilde{\varepsilon}_t - n^{-1} \sum_{k=1}^n \widetilde{\varepsilon}_k$.
3. Draw the bootstrap innovations via Efron's bootstrap from the empirical distribution function $\widehat{F}_\varepsilon(\cdot) = n^{-1} \sum_{k=1}^n \mathbb{1}_{\widehat{\varepsilon}_k \leq \cdot}$.
4. Choose some initial vector $(X_0^*, \ldots, X_{1-p}^*)'$.
5. Generate the bootstrap variables $X_t^* = \sqrt{\widehat{\theta}_{n,0} + \widehat{\theta}_{n,1}(X_{t-1}^*)^2 + \cdots + \widehat{\theta}_{n,p}(X_{t-p}^*)^2}\, \varepsilon_t^*$, $t = 1, 2, \ldots$.

Lemma 4.6. *Let $(X_t)_t$ be an ARCH(p) process with $\theta \in \Theta$, where Θ is defined as in (4.7). A bootstrap sample generated by the Algorithm [B2] meets the condition (A1*) if the probability density of the innovations is bounded and if the initial vector is drawn from the stationary bootstrap distribution, provided its existence.*

Proof. (1) To verify that the bootstrap process is stationary with probability tending to one, we first note that the inequalities $\sum_{i=1}^p \widehat{\theta}_{n,i} < 1$ and $\widehat{\theta}_{n,0} > 0$ hold with probability tending to one. According to their construction, the bootstrap innovations are standardized. Therefore, it exists a stationary solution to the model equation of [B2] 5. with probability tending to one. Finally, starting the process with its stationary distribution, provided it exists, yields the desired stationarity property of the bootstrap process.

(2) As in the case of AR(p) bootstrap, convergence of the two-dimensional distributions with probability tending to one follows from Corollary 4.1 of Neumann and Paparoditis [94] in conjunction with their Lemmas 4.3, where they assumed the innovation density to be bounded.

(3) Define $(\mathfrak{X}_n)_{n\in\mathbb{N}}$, such that $\widehat{\theta}_n(\mathbb{X}_n = x_n) \in \Theta$ with $\|\widehat{\theta}_n - \theta\|_1 \leq \delta_n$ for a null sequence $(\delta_n)_n$ and all $(x_n)_n$ with $x_n \in \mathfrak{X}_n,\ n \in \mathbb{N}$. Due to our preliminary considerations, this ensures the existence of a unique stationary distribution (conditionally on $X_1, \ldots, X_n$). Moreover, the sets can be chosen in such a way that $P(\mathbb{X}_n \in \mathfrak{X}_n) \longrightarrow_{n\to\infty} 1$. Let the initial vector $(X_0^*, \ldots, X_{1-p}^*)'$ be then drawn from the stationary distribution. Since the innovations are standardized, we get $\mathbb{E}((X_1^*)^2 \mid \mathbb{X}_n = x_n) \leq C$ for all $(x_n)_n$ with $x_n \in \mathfrak{X}_n,\ n \in \mathbb{N}$. Applying the same coupling method as in the proof of Lemma 2.1 of Neumann and Paparoditis [93], one obtains an exponential decay of the elements of the sequence $(\bar{\tau}_r)_r$.

$\square$

4.4.3 Sieve bootstrap for linear processes

We consider the bootstrap method proposed by Bühlmann [21] that is described in Subsection 4.3.1.

Lemma 4.7. *Suppose that the assumptions of Lemma 4.4 are satisfied. Moreover, let* $p(n) = o((n/log(n))^{1/(2l+2)})$ *and assume that* $\sum_{k=1}^{\infty} k^l |a_k| < \infty$ *for some* $l \geq 7,\ l \in \mathbb{N}$. *Then the bootstrap algorithm of Bühlmann [21] satisfies the condition* $(A1^*)$ *with* $\delta \in [1/m, 1)$, *where* $m < l/3,\ m \in \mathbb{N}$.

Proof. (1) Since Yule-Walker estimators are used to determine the coefficients of the bootstrap process, stationarity is preserved. For details we refer the reader to Brockwell and Davis [20], Section 8.1.

(2) Convergence of the fidis is assured by Lemma 4.4.

(3) Obviously, it suffices to consider the case $\delta = 1/m$. Under our assumptions the assertion of Lemma 5.1 of Bühlmann [21] holds true, i.e. there exist a random variable N and a finite constant K such that $\sup_{n \geq N} \sum_{k=1}^{\infty} k^l |\widehat{a}_{n,k}| \leq K$ almost surely. Here, $(\widehat{a}_{n,k})_k$ denotes the sequence of coefficients in the MA(∞) representation of the bootstrap process conditionally on $X_1, \dots, X_n$. We specify $(\mathfrak{X}_n^{(1)})_n$, introduced in $(A1^*)$, as follows

$$\mathfrak{X}_n^{(1)} := \left\{ x_n \Big| \sum_{k=1}^{\infty} k^l |\widehat{a}_{n,k}| \leq K,\ \mathbb{E}(|\varepsilon_1^*| \mid \mathbb{X}_n = x_n) \leq 2\,\mathbb{E}|\varepsilon_1| \right\},$$

which implies $P(\mathbb{X}_n \in \mathfrak{X}_n^{(1)}) \longrightarrow_{n \to \infty} 1$ by virtue of Lemma 5.3 of Bühlmann [21]. Now, let $(x_n)_n$ be an arbitrary sequence with $x_n \in \mathfrak{X}_n^{(1)},\ n \in \mathbb{N}$. According to Example 2.1.3 in Subsection 2.1.2, we have $\tau_r^*(x_n) \leq 6\,\mathbb{E}(|\varepsilon_1^*| \mid \mathbb{X}_n = x_n) \sum_{k=r}^{\infty} |\widehat{a}_{n,k}(x_n)|,\ r \in \mathbb{N}$. Thus, it remains to show that $\sum_{r=1}^{\infty} r\,(\bar{\tau}_r)^{\delta} < \infty$ with

$$\bar{\tau}_r = 12\,\mathbb{E}|\varepsilon_1| \sup_{n \in \mathbb{N}} \sup_{x_n \in \mathfrak{X}_n} \sum_{k=r}^{\infty} |\widehat{a}_{n,k}(x_n)|.$$

One can estimate

$$\begin{aligned} \sum_{r=1}^{\infty} r\,\bar{\tau}_r^{\delta} &\leq C \sum_{r=1}^{\infty} r \left(\sum_{k=r}^{\infty} \sup_{n \in \mathbb{N}} \sup_{x_n \in \mathfrak{X}_n} |\widehat{a}_{n,k}(x_n)| \right)^{1/m} \\ &= C \sum_{r=1}^{\infty} r \lim_{R \to \infty} \left(\sum_{k=r}^{R} \sup_{n \in \mathbb{N}} \sup_{x_n \in \mathfrak{X}_n} |\widehat{a}_{n,k}(x_n)| \right)^{1/m}. \end{aligned}$$

The term in brackets is less than $(\sum_{k=r}^{R} \sup_{n \in \mathbb{N}} \sup_{x_n \in \mathfrak{X}_n^{(1)}} |\widehat{a}_{n,k}(x_n)|^{1/m})^m$ since $(a+b) \leq (a^{1/m} + b^{1/m})^m$ for $a, b \geq 0$ according to the Binomial theorem. Hence, interchanging the order of summation yields

$$\begin{aligned}\sum_{r=1}^{\infty} r(\bar{\tau}_r)^{\delta} &\leq C\sum_{r=1}^{\infty} r\left(\sum_{k=r}^{\infty} \sup_{n\in\mathbb{N}} \sup_{x_n\in\mathfrak{X}_n^{(1)}} |\widehat{a}_{n,k}(x_n)|^{1/m}\right)\\ &\leq C\sum_{r=1}^{\infty} \frac{r(r+1)}{2} \sup_{n\in\mathbb{N}} \sup_{x_n\in\mathfrak{X}_n^{(1)}} |\widehat{a}_{n,r}(x_n)|^{1/m}\\ &\leq C\sum_{r=1}^{\infty} r^2\, r^{-l/m},\end{aligned}$$

which is finite according to the choice of the numbers l and m.

□

4.4.4 Bootstrapping contractive iterated random functions

Numerous (non-)linear time series models are expressed in terms of iterative random functions, that is, they have a Markovian representation

$$X_k = G(X_{k-1}, \ldots, X_{k-p}, \varepsilon_k), \quad k \in \mathbb{Z}. \tag{4.8}$$

Here, the process $(X_k)_{k\in\mathbb{Z}}$ has values in $\mathbb{R}^d$ and $(\varepsilon_k)_{k\in\mathbb{Z}}$ is a sequence of i.i.d. $\mathbb{R}^q$-valued random variables. Besides ordinary AR(p) processes and ARCH(p) processes, examples include nonlinear autoregressive processes such as exponential autoregressive models (EXPAR(p)) and functional autoregressive models with exogenous random variables, i.e.

$$Y_k = g(Y_{k-1}, \ldots, Y_{k-p_1}, Z_k, Z_{k-1}, \ldots, Z_{k-p_2}) + \eta_k, \quad k \in \mathbb{Z}.$$

The $\mathbb{R}^d$-valued innovations $(\eta_k)_k$ are assumed to be i.i.d. and $(Z_k)_{k\in\mathbb{Z}}$ denotes a sequence of i.i.d. exogenous variables with values in $\mathbb{R}^m$. Defining $X_k = (Y_k', Z_k')'$ and

$$G(X_{k-1}, \ldots, X_{k-\max\{p_1,p_2\}}, \varepsilon_k) = \begin{pmatrix} g(Y_{k-1}, \ldots, Y_{k-p_1}, Z_k, Z_{k-1}, \ldots, Z_{k-p_2}) + \eta_k \\ Z_k \end{pmatrix}$$

with $\varepsilon_k = (\eta_k', Z_k')'$, this process can be reformulated in terms of (4.8). Recall that stationarity of this kind of process is assured if it is contracting on average. More precisely, the following two conditions are sufficient:

(I) $\exists\, y_0 \in \mathbb{R}^{dp}$ with $\mathbb{E}\|G(y_0, \varepsilon_0)\|_2 < \infty$,

(II) $\exists\, a_1, \ldots, a_p \geq 0$ with $\sum_{k=1}^{p} a_k < 1$ and

$$\mathbb{E}\|G(y, \varepsilon_0) - G(\bar{y}, \varepsilon_0)\|_2 \leq \sum_{k=1}^{p} a_k \|y_k - \bar{y}_k\|_2, \quad \forall\, y = (y_1', \ldots, y_p')',\ \bar{y} = (\bar{y}_1', \ldots, \bar{y}_p')'.$$

According to Theorem 5.1 of Shao and Wu [102], the sequence $(X_k)_k$ is τ-dependent with exponentially decaying coefficients under the assumptions (I) and (II).

Based on Lemma 4.3 we verify the validity of condition (A1*) for the following parametric bootstrap method that imitates $X_k = G(X_{k-1}, \dots, X_{k-p}, \varepsilon_k; \theta_0), \quad k \in \mathbb{Z},\ \theta_0 \in \Theta \subseteq \mathbb{R}^m$.

[B3] Algorithm.

1. Calculate the parameter estimator $\widehat{\theta}_n(X_1, \dots, X_n)$ such that $\widehat{\theta}_n \xrightarrow{P} \theta_0$.
2. Draw i.i.d. bootstrap innovations $\varepsilon_k^*,\ k \geq 1$, such that $\varepsilon_1^* \xrightarrow{d} \varepsilon_1$, in probability.
3. Choose some initial vector $(X_0^{*\prime}, \dots, X_{1-p}^{*\prime})'$.
4. Generate $X_k^* = G(X_{k-1}^*, \dots, X_{k-p}^*, \varepsilon_k^*; \widehat{\theta}_n),\ k = 1, 2, \dots$.

Lemma 4.8. *Let $(X_k)_{k\in\mathbb{Z}}$ be the stationary solution of $X_k = G(X_{k-1}, \dots, X_{k-p}, \varepsilon_k; \theta_0), \quad k \in \mathbb{Z},\ \theta_0 \in \Theta$, satisfying the conditions (I) and (II), where Θ is an open subset of $\mathbb{R}^m$. Moreover, assume the function $G : \mathbb{R}^{dp+q} \times \Theta \to \mathbb{R}^d$ to be continuous. Suppose that the following conditions are fulfilled:*

(a) There exists a $y_0^* \in \mathbb{R}^d$ and some $K_1 \in \mathbb{R}$ such that $\mathbb{E}^* \|G(y_0^*, \varepsilon_0^*; \widehat{\theta}_n)\|_2 \leq K_1$, with probability tending to one.

(b) There are a constant $\delta \in (0,1)$ and nonnegative random variables $a_{n,1}, \dots, a_{n,p}$ with $P(\sum_{i=1}^p a_{n,i} \leq 1 - \delta) \longrightarrow_{n\to\infty} 1$ and such that

$$\mathbb{E}^* \|G(y, \varepsilon_0^*; \widehat{\theta}_n) - G(\bar{y}, \varepsilon_0^*; \widehat{\theta}_n)\|_2 \leq \sum_{k=1}^p a_{n,k} \|y_k - \bar{y}_k\|_2, \quad \forall\, y, \bar{y} \in \mathbb{R}^{dp},$$

Then the bootstrap process generated by Algorithm [B3] satisfies condition (A1) if the initial vector is drawn from the stationary bootstrap distribution, provided its existence.*

Proof. (1) Define

$$\mathfrak{X}_n := \Big\{ \mathbb{X}_n = x_n \Big| \sum_{i=1}^p |a_{n,i}| \leq 1 - \delta,\ \mathbb{E} \left(\|G(y_0^*, \varepsilon_0^*; \widehat{\theta}_n)\|_2 \mid \mathbb{X}_n = x_n \right) \leq K_1, \\ \|\widehat{\theta}_n - \theta_0\|_1 \leq \delta_n,\ \widehat{\theta}_n \in \Theta \Big\},\ n \in \mathbb{N},$$

where the null sequence $(\delta_n)_n$ can be chosen such that $P(\mathbb{X}_n \in \mathfrak{X}_n) \longrightarrow_{n\to\infty} 1$. Thus, there exists a stationary law to the bootstrap model provided $\mathbb{X}_n \in \mathfrak{X}_n$, cf. Example 2.2.1 of Subsection 2.1.2. Then, drawing $X_0^*, \dots, X_{1-p}^*$ from this distribution, we obtain (1).

(2) In order to verify convergence of the fidis, Lemma 4.3 is invoked. Thus, we have to check whether its prerequisites (i) to (iv) hold in probability. Condition (i) is satisfied if for each $\delta > 0$ and each compact subset K of $\mathbb{R}^{dp}$,

$$P\left(\sup_{y\in K} d\left(P_{G(y_1,\ldots,y_p,\varepsilon_1^*;\widehat{\theta}_n)|\mathbb{X}_n}, P_{G(y_1,\ldots,y_p,\varepsilon_1;\theta_0)}\right) > \delta\right) \xrightarrow{P} 0.$$

To this end, suppose that the bootstrap is started with its stationary distribution if it exists and define $\widetilde{\mathfrak{X}}_n \subseteq \mathfrak{X}_n$ such that moreover $P_{\varepsilon_1^*|\mathbb{X}_n=x_n} \Longrightarrow P_{\varepsilon_1}$ uniformly for all sequences $(x_n)_n$ with $x_n \in \widetilde{\mathfrak{X}}_n,\ n \in \mathbb{N}$, and $P(\mathbb{X}_n \in \widetilde{\mathfrak{X}}_n) \longrightarrow_{n\to\infty} 1$. We construct a grid $y^{(1)},\ldots,y^{(M)}$ in K such that $\sup_{y\in K}\min\{\|y-y^{(i)}\|_2 \mid i\in\{1,\ldots,M\}\} \le \delta/3$. Let $(x_n)_n$ be an arbitrary sequence with $x_n \in \widetilde{\mathfrak{X}}_n,\ n\in\mathbb{N}$. The application of the triangular inequality yields

$$\begin{aligned}
&\sup_{y\in K} d\left(P_{G(y_1,\ldots,y_p,\varepsilon_1^*;\widehat{\theta}_n)|\mathbb{X}_n=x_n}, P_{G(y_1,\ldots,y_p,\varepsilon;\theta_0)}\right)\\
&\le \frac{\delta}{3}\sum_{i=1}^{p}[a_k + a_{n,k}(x_n)] + \sum_{i=1}^{M} d\left(P_{G(y_1^{(i)},\ldots,y_p^{(i)},\varepsilon_1^*;\widehat{\theta}_n)|\mathbb{X}_n=x_n}, P_{G(y_1^{(i)},\ldots,y_p^{(i)},\varepsilon_1;\theta_0)}\right),
\end{aligned}$$

where the first summand can be bounded from above by $2\,\delta/3$. The second one is less than $\delta/3$ for all $n \ge n_0$ for some $n_0 \in \mathbb{N}$ since G is continuous and $P_{(\varepsilon_1^{*\prime},\widehat{\theta}_n')'|\mathbb{X}_n=x_n} \Longrightarrow P_{(\varepsilon_1',\theta_0')'}$ for all $(x_n)_n$ with $x_n \in \widetilde{\mathfrak{X}}_n,\ n\in\mathbb{N}$. This implies the condition (i) of Lemma 4.3.

Regarding the constraint (ii) we obtain

$$\begin{aligned}
&\sup_{\widetilde{y}:\|\widetilde{y}-y\|_2\le\delta} d\left(P_{Y_t|Y_{t-1}=y_1,\ldots,Y_{t-p}=y_p}, P_{Y_t|Y_{t-1}=\widetilde{y}_1,\ldots,Y_{t-p}=\widetilde{y}_p}\right)\\
&\le \sup_{\widetilde{y}:\|\widetilde{y}-y\|_2\le\delta} \mathbb{E}\|G(y,\varepsilon_0;\theta_0) - G(\widetilde{y},\varepsilon_0;\theta_0)\|_2\\
&\le \sup_{\widetilde{y}:\|\widetilde{y}-y\|_2\le\delta} \sum_{k=1}^{p} a_k\|y_k - \bar{y}_k\|_2,
\end{aligned}$$

where the latter expression tends to zero as $\delta \to 0$.

Next, tightness of the bootstrap process has to be verified. To this end, we consider again an arbitrary sequence $(x_n)_n$ with $x_n \in \mathfrak{X}_n,\ n\in\mathbb{N}$. Based on $\mathbb{X}_n = x_n$, let $(X_k^*)_k$ be the bootstrap process that is started with its stationary distribution. In order to prove tightness of the bootstrap process, it now suffices to show that $\sup_n \mathbb{E}(\|X_k^*\|_2 \mid \mathbb{X}_n = x_n) \le K_2$ for some finite constant K_2 and all sequences $(x_n)_n$ with $x_n \in \mathfrak{X}_n,\ n\in\mathbb{N}$. According to Example 2.2.1, $\sup_{k\in\mathbb{Z}}\mathbb{E}(\|X_k^*\|_2 \mid \mathbb{X}_n = x_n)$ is finite. We obtain

$$\begin{aligned}
&\mathbb{E}(\|X_k^*\|_2 \mid \mathbb{X}_n = x_n)\\
&\le \mathbb{E}(\|G(X_{k-1}^*,\ldots,X_{k-p}^*,\varepsilon_k^*;\widehat{\theta}_n) - G(y_0^*,\varepsilon_k^*;\widehat{\theta}_n)\|_2 \mid \mathbb{X}_n = x_n) + K_1\\
&\le \sum_{j=1}^{p} a_{n,j}\mathbb{E}(\|X_{k-1}^*\|_2 \mid \mathbb{X}_n = x_n) + \|y_0^*\|_2 + K_1.
\end{aligned}$$

By stationarity this implies

$$\mathbb{E}(\|X_k^*\|_2 \mid \mathbb{X}_n = x_n) \leq \frac{\|y_0^*\|_2 + K_1}{1 - \sum_{j=1}^p a_{n,j}} \leq \frac{\|y_0^*\|_2 + K_1}{\delta}, \quad k \in \mathbb{N},$$

for any sequence $(x_n)_n$ with $x_n \in \mathfrak{X}_n,\ n \in \mathbb{N}$.

Finally, the condition (iv) of Lemma 4.3 follows from Theorem 5.1 of Shao and Wu [102] since they show existence of a stationary distribution and convergence of every arbitrarily started process towards this limit. This in turn yields its uniqueness. Consequently, for all $k \in \mathbb{N}$, $P_{X_1^*,\dots,X_k^*} \Longrightarrow P_{X_1,\dots,X_k}$, in probability.

(3) Let $(x_n)_n$ be an arbitrary sequence with $x_n \in \mathfrak{X}_n,\ n \in \mathbb{N}$. Moreover, draw two independent initial values $X_0^*,\dots,X_{1-p}^*$ and $\widetilde{X}_0^*,\dots,\widetilde{X}_{1-p}^*$ for the bootstrap process from its stationary distribution and define $\widetilde{X}_k^* = G(\widetilde{X}_{k-1}^*,\dots,\widetilde{X}_{k-p}^*,\varepsilon_k^*,\widehat{\theta}_n),\ k \in \mathbb{N}$. The bootstrap process is τ-dependent with exponentially decaying coefficients if $\mathbb{E}(\|X_r^* - \widetilde{X}_r^*\|_1 \mid \mathbb{X}_n = x_n) \leq C\,(\bar{\tau})^r$ with some $\bar{\tau} \in (0,1)$. According to the proofs of Proposition 6.3.22 and Lemma 6.2.10 of Duflo [54], this inequality is fulfilled whenever the the spectral radii of the matrices

$$A_n = \begin{pmatrix} a_{n,1} & a_{n,2} & \cdots & a_{n,p-1} & a_{n,p} \\ 1 & 0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}, \quad n \in \mathbb{N},$$

are less than a constant that is strictly smaller than one uniformly on $(\mathfrak{X}_n)_n$. By virtue of Lemma 4.1.1 of Duflo [54], the spectral radius is the absolute largest inverse of the zeros of the polynomial $1 - \sum_{k=1}^n a_{n,k}\,z^k,\ z \in \mathbb{C}$. We obtain

$$\left|1 - \sum_{k=1}^n a_{n,k}\,z^k\right| \geq \left|1 - \sum_{k=1}^p a_{n,k}|z|^k\right|.$$

Thus, the modulus of any zero of the polynomial above is greater than one, which in turn implies

$$1 - \sum_{k=1}^n a_{n,k}|z|^k \geq 1 - |z|^p \sum_{k=1}^p a_{n,k} \geq 1 - |z|^p(1-\delta)$$

due to the construction of the sets $\mathfrak{X}_n$. Therefore, a necessary condition for $z \in \mathbb{C}$ being a zero of the considered polynomial is given by $|z| \geq 1/(1-\delta)$. This leads to an upper bound $(1-\delta)$ for the spectral radius of $A_n,\ n \in \mathbb{N}$.

□

5 Application to hypothesis tests of L_2-type

5.1 Overview

During the last decades numerous time series models have been intensively applied to describe various real-life phenomena, e.g. in finance, physics, and biometrics. Based on the results of the previous two chapters, we provide consistent testing procedures that allow for checking the adequacy of certain models.

Let $X_1, \ldots, X_n$ be $\mathbb{R}^d$-valued observations of a stationary weakly dependent process with marginal distribution P_X. We consider problems of the following structure:

$$\mathcal{H}_0: \quad g_X(\cdot) = f_X(\cdot, \theta_0) \quad \text{a.s. for some } \theta_0 \in \Theta \subseteq \mathbb{R}^p \quad \text{against}$$
$$\mathcal{H}_1: \quad g_X(\cdot) \neq f_X(\cdot, \theta) \quad \text{on a set of positive measure for all } \theta \in \Theta.$$

Besides supremum-type tests, L_2-tests are the most convenient in mathematical statistics. Here, test statistics of L_2-type $\widehat{T}_n = n \int_{\mathbb{R}^m} [\psi(g_X^{(n)}(t) - f_X(t, \widehat{\theta}_n), t, \widehat{\theta}_n)]^2 \, dt$ are applied, where $\widehat{\theta}_n$ denotes a parameter estimator. The functions f_X, g_X, ψ, as well as $g_X^{(n)}$, which denotes an appropriate fixed-kernel non-parametric estimate of g_X, are specified in the Sections 5.3 to 5.5. There are two basic approaches in order to derive the asymptotic null distribution of the statistic $\widehat{T}_n$. One can either invoke empirical process theory or apply asymptotic results on degenerate U-statistics. In this thesis, the latter method is utilized.

In Section 5.3 we extend the symmetry test which was initially proposed by Feuerverger and Mureika [67] for i.i.d. data and a known center of symmetry to the case of weakly dependent observations with an unknown location parameter. Answering the question whether a distribution is symmetric or not is interesting for several reasons. First, there is a pure statistical interest since the presence or absence of symmetry is important for deciding what parameter to estimate. If the underlying distribution is symmetric, the point of symmetry is the only reasonable location parameter, whereas in the nonsymmetric case there is no longer only one measure of location, cf. Antille, Kersting and Zucchini [4] and references therein. Moreover, symmetry plays a central role in analyzing and modeling business circles, see Ramsey and Rothman [98] for a detailed discussion of this topic. Finally, rejecting the hypothesis of symmetry has substantial impact on model selection. In this case, fitting the data to a nonlinear AR(p) process with a skew-symmetric regression function and symmetric innovations is inappropriate, cf. Pemberton and Tong [96].

In Section 5.4 a goodness-of-fit test for the marginal distribution of a time series is constructed. So far, the normality assumption is still dominating the literature in many fields, however, not for reason of empirical evidence but for theoretical simplicity. Within the last decades the literature on non-Gaussian time series has addressed the topic of finding innovation distributions that lead to specified marginals of a time series model. Surveys of those results are given by Jose, Tomy and Sreekumar [78] as well as by Block, Langberg and Stoffer [17]. Goodness-of-fit tests for dependent data based on the L_2-distance between the non-parametric kernel density estimate with vanishing bandwidth and a smoothed version of a parametric estimate of the density were considered by Fan and Ullah [66] as well as by Neumann and Paparoditis [92]. The corresponding test statistics are asymptotically normal. Here, we propose an L_2-test that evaluates the difference between the characteristic function and its fixed-kernel estimate, the empirical characteristic function. The test statistic does not behave asymptotically normal but can be approximated by a degenerate V-statistic. Thus our results of Chapter 3 and Chapter 4 can be invoked to investigate the asymptotics of the test statistic.

In Section 5.5 fixed-kernel L_2-tests for parametric models of the conditional mean function are established. An overwhelming amount of the statistical and econometrics literature is dedicated to consistent model specification tests. An extensive list of tests in the i.i.d. context is provided by Escanciano [60]. In the time series framework, kernel-based tests with vanishing bandwidth were for instance considered by Fan and Li [63], Hjellvik, Yao and Tjøstheim [72], Fan and Li [64], as well as by Kreiss, Neumann and Yao [82]. A second class of tests extends the so-called integrated conditional moment test of Bierens [14] for i.i.d. data towards dependent observations. Examples include the work of de Jong [46], Bierens and Ploberger [15], Koul and Stute [81], as well as of Escanciano [60]. Since the asymptotic null distribution depends on unknown parameters, Escanciano [60] proposed the application of wild bootstrap procedures in order to derive critical values of the test. Here, we will justify a parametric bootstrap procedure. The latter method may perform better since, in contrast to the wild bootstrap, the bootstrap counterparts of the observed random variables converge in distribution to original ones with probability tending one.

In all three cases we do not only investigate the behaviour of our tests under $\mathcal{H}_0$ and $\mathcal{H}_1$ but also verify asymptotic unbiasedness against certain local alternatives. To this end, auxiliary results regarding U-statistics under contiguous alternatives and U-statistics with varying kernels, respectively, are required. These are established in the subsequent section.

5.2 Some auxiliary results

5.2.1 Auxiliary results on U- and V-statistics for triangular schemes of random variables

In order to develop the behaviour under local alternatives of the tests that are established in Section 5.3 and Section 5.4, additional asymptotic results concerning U- and V-statistics for triangular schemes of random variables are required. In Chapter 3 we developed the limit distributions of U- and V-statistics for stationary sequences of random variables with the property $\int_{\mathbb{R}^d} h(x,y)P_{X_1}(dx) = 0,\ \forall\, y \in \mathbb{R}^d$. These results were extended to row-wise stationary triangular schemes of random random variables in Chapter 4. Doing so, the assumption of degeneracy, i.e. $\int_{\mathbb{R}^d} h(x,y)P_{X_1^*|\mathbb{X}_n}(dx) = 0,\ \forall\, y \in \mathbb{R}^d$, has been maintained. Now, we derive the asymptotics of U- and V-statistics for row-wise stationary triangular schemes of random variables where we relax the latter constraint. More precisely, we make the following assumptions concerning the underlying random variables:

(A5) **(i)** The triangular scheme $(X_{n,k})_{k=1}^n,\ n \in \mathbb{N}$, of row-wise stationary and $\mathbb{R}^d$-valued random variables on a probability space $(\Omega, \mathcal{A}, P)$ satisfies $\sup_{n\in\mathbb{N}} \mathbb{E}\|X_{n,1}\|_1 \leq C_1$ for some $C_1 < \infty$. Moreover, the sequence $(\bar{\tau}_r)_{r\in\mathbb{N}}$, defined as $\bar{\tau}_r := \sup_{n>r} \tau_{r,n}$, where

$$\tau_{r,n} := \sup\{\tau(\sigma(X_{n,s_1}, \ldots, X_{n,s_u}), (X'_{n,t_1}, X'_{n,t_2}, X'_{n,t_3})') \mid \\ u \in \mathbb{N}, 1 \leq s_1 \leq \cdots \leq s_u < s_u + r \leq t_1 \leq t_2 \leq t_3 \leq n\},$$

fulfils $\sum_{r=1}^{\infty} r\,(\bar{\tau}_r)^{\delta^2} < \infty$ for some $\delta \in (0,1)$.

(ii) There exists a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ satisfying the assumption (A1) with $\sum_{r=1}^{\infty} r\,(\tau_r)^{\delta^2} < \infty$ and such that $\left(X'_{n,t_1}, X'_{n,t_2}\right)' \xrightarrow{d} \left(X'_{t_1}, X'_{t_2}\right)',\ 1 \leq t_1, t_2 \leq n,\ n \in \mathbb{N}$. The Radon Nikodym derivatives are given by $dP_{X_{n,1}}/dP_{X_1} = 1 + n^{-1/2} g_n$, where $(g_n)_{n\in\mathbb{N}}$ is a sequence of functions with $\|g_n - g\|_\infty \longrightarrow_{n\to\infty} 0$ for some bounded measurable function g.

Example 5.1. Let $(X_n)_{n\in\mathbb{N}}$ be a τ-dependent, stationary, real-valued process with $\sum_{r=1}^{\infty} r\,(\tau_r)^{\delta^2} < \infty$ for some $\delta \in (0,1)$ and such that P_{X_1} has a twice continuously differentiable Lebesgue density with $\|f\|_\infty + \|f^{(1)}\|_\infty + \|f^{(2)}\|_\infty < \infty$. Then a process that satisfies the condition (A5) can be easily constructed: Consider a sequence of i.i.d. real-valued, non-centered and square integrable random variables $(Z_n)_{n\in\mathbb{N}}$ that is independent of $(X_n)_{n\in\mathbb{N}}$. Now define $X_{n,k} := X_k + n^{-1/2} Z_k$. Obviously, the triangular scheme $(X_{n,k})_{k=1}^n,\ n \in \mathbb{N}$, is row-wise stationary and satisfies the dependence condition above. In order to verify convergence of the two-dimensional distributions, let $G : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a bounded Lipschitz continuous function. We obtain

$$|\mathbb{E}G(X_{n,t_1}, X_{n,t_2}) - \mathbb{E}G(X_{t_1}, X_{t_2})| \leq \mathrm{Lip}(G)\,(\mathbb{E}|X_{n,t_1} - X_{t_1}| + \mathbb{E}|X_{n,t_2} - X_{t_2}|) \\ \leq \frac{C}{\sqrt{n}}\,\mathbb{E}|Z_1| \underset{\mathrm{n}\to\infty}{\longrightarrow} 0.$$

Finally, concerning the Radon-Nikodym derivatives, Taylor expansion yields

$$f_{X_{n,1}}(x) = \mathbb{E} f_{X_1}\left(x - \frac{1}{\sqrt{n}} Z_1\right) = f_{X_1}(x) - \frac{1}{\sqrt{n}}\, f_{X_1}^{(1)}(x)\,\mathbb{E}Z_1 + \frac{1}{2n}\,\mathbb{E} f_{X_1}^{(2)}(\xi)\, Z_1^2$$

for some random ξ between $x - Z_1/\sqrt{n}$ and x, and thus the desired result with $g_n(x) = -f_{X_1}^{(1)}(x)\,\mathbb{E}Z_1 + [2\sqrt{n}]^{-1}\,\mathbb{E} f_{X_1}^{(2)}(\xi)\, Z_1^2$.

Regarding the kernel functions we assume:

(A6) **(i)** The kernel $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a symmetric, measurable function and degenerate under P_{X_1}, where X_1 is defined as in (A5).

(ii) For a δ satisfying (A5), the following moment constraints hold true with some $\nu > (2-\delta)/(1-\delta)$:

$$\sup_{\substack{1\le k<n,\\ n\in\mathbb{N}}} \mathbb{E}|h(X_{n,1}, X_{n,1+k})|^\nu < \infty \quad \text{and} \quad \mathbb{E}|h(X_1, \widetilde{X}_1)|^\nu < \infty,$$

where $\widetilde{X}_1$ denotes an independent copy of X_1.

(iii) The kernel function satisfies

$$|h(x,y) - h(\bar{x},\bar{y})| \le f(x,\bar{x},y,\bar{y})\,[\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1]\,,\ \forall\, x,\, \bar{x},\, y,\, \bar{y} \in \mathbb{R}^d,$$

for a continuous function $f : \mathbb{R}^{4d} \to \mathbb{R}$ that is symmetric under permutations of its arguments. Moreover, let $\delta \in (0,1)$ such that (A6)(ii) holds. Then

$$\sup_{\substack{1\le k_1,\ldots,k_5\le n,\\ n\in\mathbb{N}}} \mathbb{E}\Big(\max_{a_1,a_2\in[-A,A]^d} [f(Y_{n,k_1}, Y_{n,k_2}+a_1, Y_{n,k_3}, Y_{n,k_4}+a_2)]^\eta \|Y_{n,k_5}\|_1 \Big)$$

is finite for $\eta := 1/(1-\delta)$, some $A > 0$, and any $(Y'_{n,k_1},\ldots,Y'_{n,k_5})'$ consisting of independent subvectors $(Y'_{n,k_{j_1(m)}},\ldots,Y'_{n,k_{j_l(m)}})'$, $l,m = 1,\ldots,5$, where either $(Y'_{n,k_{j_1(m)}},\ldots,Y'_{n,k_{j_l(m)}})' \overset{d}{=} (X'_{n,k_{j_1(m)}},\ldots,X'_{n,k_{j_l(m)}})'$ or $(Y'_{n,k_{j_1(m)}},\ldots,Y'_{n,k_{j_l(m)}})' \overset{d}{=} (X'_{k_{j_1(m)}},\ldots,X'_{k_{j_l(m)}})'$.

The conditions listed in (A5) basically coincide with those of (A1*). The additional assumption concerning the Radon-Nikodym derivatives is required to compensate the lack of degeneracy of the kernel under the distribution $P_{X_{n,1}}$. Note that assumption (A6)(i) merely states degeneracy under P_{X_1} but not under the sequence of laws $(P_{X_{n,1}})_n$. In conjunction with condition (A5)(ii) we obtain asymptotic degeneracy in the following sense:

$$\left|\int_{\mathbb{R}^d} H(x,y) P_{X_{n,1}}(dx) P_{X_{n,1}}(dy)\right| \le n^{-1}\,\|g_n\|_\infty^2\, \mathbb{E}\big|H(X_1,\widetilde{X}_1)\big| \underset{\mathrm{n}\to\infty}{\longrightarrow} 0$$

for *any* symmetric kernel H that degenerates under P_{X_1}. This turns out to be crucial for proving the asymptotic results below. Comparing the remaining part of the assumption (A6) to the constraints (A2*) and (A3*), respectively, we observe that they are almost identical. Again, by means of the assumptions concerning the Radon-Nikodym derivatives, the condition $\mathbb{E}|h(X_1,\widetilde{X}_1)|^\nu < \infty$ implies $\mathbb{E}|h(X_{n,1},\widetilde{X}_{n,1})|^\nu < \infty$ for all sufficiently large n, which would turn (A6)(ii) into the exact counterpart of the assumption (A2*)(iii). Moreover, note that the conditions (A5) and (A6) imply validity of the assumptions (A2)(i) and (A4)(i) for the sample $(X_k)_k$ defined in (A5). Finally, we remark that the summability condition concerning the τ-dependence coefficients in (A5) can be weakened to $\sum_{r=1}^\infty \bar{\tau}_r^\delta < \infty$ and $\sum_{r=1}^\infty \tau_r^\delta < \infty$, respectively, if the function h is Lipschitz continuous.

Based on these prerequisites, the limit distributions of

$$n\,U_{n,n} := \frac{1}{n(n-1)} \sum_{\substack{j,k=1\\ j\neq k}}^{n} h(X_{n,j}, X_{n,k}) \quad \text{and} \quad n\,V_{n,n} := \frac{1}{n^2} \sum_{j,k=1}^{n} h(X_{n,j}, X_{n,k}).$$

can be established.

Proposition 5.1. *Suppose that the assumptions (A5) and (A6) are fulfilled. Then, as $n \to \infty$,*

$$n\,U_{n,n} \xrightarrow{d} Z_{loc},$$

where

$$\begin{aligned} Z_{loc} := \lim_{c\to\infty} \Bigg\{ & \sum_{k_1,k_2\in\mathbb{Z}^d} \alpha^{(c)}_{k_1,k_2} [(Z_{k_1} + C_{k_1})(Z_{k_2} + C_{k_2}) - A_{k_1,k_2}] \\ & + \sum_{j=0}^{\infty} \sum_{k_1,k_2\in\mathbb{Z}^d} \sum_{e=(e_1',e_2')'\in\bar{E}} \beta^{(c,e)}_{j;k_1,k_2} \left[\left(Z^{(e_1)}_{j;k_1} + D^{(e_1)}_{j;k_1}\right) \left(Z^{(e_2)}_{j;k_2} + D^{(e_2)}_{j;k_2}\right) - B^{(e)}_{j;k_1,k_2} \right] \Bigg\}. \end{aligned} \tag{5.1}$$

Here, the quantities $\alpha^{(c)}_{k_1,k_2}$, A_{k_1,k_2}, $\beta^{(c,e)}_{j;k_1,k_2}$, $B^{(e)}_{j;k_1,k_2}$ as well as the covariance structure of the centered jointly normal variables $(Z_k)_k$ and $(Z^{(e_1)}_{j;k})_{j,k,e_1}$ are defined as in Lemma 3.6. Moreover, $C_k := \int_{\mathbb{R}^d} [\Phi_{0,k}(x) - \mathbb{E}\Phi_{0,k}(X_1)]\, g(x)\, P_{X_1}(dx)$ and $D^{(e_1)}_{j;k} := \int_{\mathbb{R}^d} [\Psi^{(e_1)}_{j,k}(x) - \mathbb{E}\Psi^{(e_1)}_{j,k}(X_1)]\, g(x)\, P_{X_1}(dx)$, $j \in \mathbb{N}_0$, $k, k_1, k_2 \in \mathbb{Z}^d$, $e_1 \in \{0,1\}^d$, $e \in \bar{E}$.
If additionally, $\mathbb{E}|h(X_1,X_1)| < \infty$, then also

$$n\,V_{n,n} \xrightarrow{d} Z_{loc} + \mathbb{E}h(X_1,X_1).$$

Remark 5.1. In comparison to the limit distributions of $n\,U_n$ and $n\,V_n$, now additional constants C_k and $D^{(e)}_{j;k}$ appear in the limiting variable, i.e. the involved normal random variables are no longer centered.

Proof. We proceed analogously to the proofs of Lemma 3.8 and Theorem 3.1. W.l.o.g. let $\|g_n\|_\infty < \infty$, $\forall\, n \in \mathbb{N}$.
Step 1: Approximation by statistics with bounded kernels.
The first step contains the approximation of the U-statistic $n\,U_{n,n}$ by a statistic $n\,U_{n,n,c}$ with the bounded kernel function h_c, defined in Lemma 3.1. In order to verify that the error $n^2\,\mathbb{E}(U_{n,n} - U_{n,n,c})^2$ converges to zero with increasing truncation parameter c, we invoke the decomposition

$$n^2\,\mathbb{E}(U_{n,n} - U_{n,n,c})^2 \le 8 \sup_{1\le k<n} \mathbb{E}|H(X_{n,1}, X_{n,1+k})|^2 + \frac{8}{(n-1)^2} \sum_{r=1}^{n-1} \sum_{t=1}^{4} \widetilde{Z}^{(t)}_{n,r} \tag{5.2}$$

with $H = h - h_c$. Here,

$$\widetilde{Z}_{n,r}^{(1)} := Z_{n,r}^{(1)} + \sum_{\substack{1\le i<j;k<l;j\le l\le n\\ r:=\min\{j,k\}-i\ge l-\max\{j,k\}}} \left|\mathbb{E}H(X_{n,i},\widetilde{X}_{n,j}^{(r)})H(\widetilde{X}_{n,k}^{(r)},\widetilde{X}_{n,l}^{(r)})\right|,$$

$$\widetilde{Z}_{n,r}^{(2)} := Z_{n,r}^{(2)} + \sum_{\substack{1\le i<j;i\le k;k<l\le n\\ r:=l-\max\{j,k\}>\min\{j,k\}-i}} \left|\mathbb{E}H(X_{n,i},X_{n,j})H(X_{n,k},\widetilde{X}_{n,l}^{(r)})\right|,$$

$$\widetilde{Z}_{n,r}^{(3)} := Z_{n,r}^{(3)} + \sum_{\substack{1\le i\le k<l<j\le n\\ r:=k-i\ge j-l}} \left|\mathbb{E}H(X_{n,i},\widetilde{X}_{n,j}^{(r)})H(\widetilde{X}_{n,k}^{(r)},\widetilde{X}_{n,l}^{(r)})\right|,$$

$$\widetilde{Z}_{n,r}^{(4)} := Z_{n,r}^{(4)} + \sum_{\substack{1\le i\le k<l<j\le n\\ r:=j-l>i-k}} \left|\mathbb{E}H(X_{n,i},\widetilde{X}_{n,j}^{(r)})H(X_{n,k}X_{n,l})\right|$$

and $Z_{n,r}^{(t)},\ t=1,\dots,4$, are defined as in (3.2), where the random variables there are now substituted by the respective variables of the triangular scheme. In order to show that $\limsup_{n\to\infty}(\sup_{1\le k<n}\mathbb{E}|H(X_{n,1},X_{n,1+k})|^2 + 8\,(n-1)^{-2}\sum_{r=1}^{n-1}\sum_{t=1}^{4} Z_{n,r}^{(t)})$ tends to zero with increasing c, we follow the lines of the proof of Lemma 3.1 and Lemma 3.8, respectively. All of the approximations remain valid since, on the one hand, $\limsup_{n\to\infty} P(X_{n,1}\notin(-c,c)^d) \le P(X_1\notin(-c,c)^d) \longrightarrow_{c\to\infty} 0$ and, on the other hand, $\mathbb{E}|h(X_{n,1},\widetilde{X}_1)|^\nu<\infty$, where $\widetilde{X}_1\sim P_{X_1}$ is independent of $X_{n,1}$. Terms of the latter form appear when bounding the expressions E_1 to E_4, introduced in the proof of Lemma 3.1, due to the artificial degeneration of the truncated kernels w.r.t. P_{X_1}.

Moreover, one can show that $(n-1)^{-2}\sum_{r=1}^{n-1}(\widetilde{Z}_{n,r}^{(t)} - Z_{n,r}^{(t)}) = o_P(1),\ t=1,\dots,4$. Exemplarily, we state the calculations for $t=2$. Several cases have to be distinguished:

$$\begin{aligned}
&\frac{1}{(n-1)^2}\sum_{r=1}^{n-1}\left|\widetilde{Z}_{n,r}^{(2)} - Z_{n,r}^{(2)}\right| \\
&\le \frac{1}{(n-1)^2}\sum_{r=1}^{n-1}\sum_{\substack{1\le i<j\le k<l\le n\\ r:=l-k>j-i\\ k-j\ge j-i}} \left|\mathbb{E}H(X_{n,i},X_{n,j})H(X_{n,k},\widetilde{X}_{n,l}^{(r)})\right| \\
&\quad + \frac{1}{(n-1)^2}\sum_{r=1}^{n-1}\sum_{\substack{1\le i<j\le k<l\le n\\ r:=l-k>j-i>k-j}} \left|\mathbb{E}H(X_{n,i},X_{n,j})H(X_{n,k},\widetilde{X}_{n,l}^{(r)})\right| \\
&\quad + \frac{1}{(n-1)^2}\sum_{r=1}^{n-1}\sum_{\substack{1\le i\le k<j<l\le n\\ r:=l-j>k-i\\ j-k\ge k-i}} \left|\mathbb{E}H(X_{n,i},X_{n,j})H(X_{n,k},\widetilde{X}_{n,l}^{(r)})\right| \\
&\quad + \frac{1}{(n-1)^2}\sum_{r=1}^{n-1}\sum_{\substack{1\le i\le k<j<l\le n\\ r:=l-j>k-i>j-k}} \left|\mathbb{E}H(X_{n,i},X_{n,j})H(X_{n,k},\widetilde{X}_{n,l}^{(r)})\right|.
\end{aligned} \tag{5.3}$$

We investigate the first and the last summand only since the remaining terms can be treated in an analogous manner. To bound the first summand from above, it is useful to introduce $\widetilde{X}_{n,k} \sim P_{X_{n,k}}$ that is independent of the vector $(X'_{n,i}, X'_{n,j})'$ and such that $\mathbb{E}\|\widetilde{X}_{n,k} - X_{n,k}\|_1 \leq \bar{\tau}_{k-j}$. This procedure may require an enlargement of the underlying probability space. Now an iterative application of Hölder's inequality yields

$$\begin{aligned}
&\left|\mathbb{E}H(X_{n,i}, X_{n,j})H(X_{n,k}, \widetilde{X}_{n,l}^{(r)})\right| \\
&\leq \left|\mathbb{E}H(X_{n,i}, X_{n,j}) \left[\int_{\mathbb{R}^d} H(X_{n,k}, y) - H(\widetilde{X}_{n,k}, y)\, P_{X_{n,1}}(dy)\right]\right| \\
&\quad + \left|\mathbb{E}H(X_{n,i}, X_{n,j})\, \mathbb{E}H(X_{n,1}, \widetilde{X}_{n,1})\right| \\
&\leq C\, \bar{\tau}_{k-j}^{\delta^2} \left(\mathbb{E}|H(X_{n,i}, X_{n,j})|^{(2-\delta)/(1-\delta)}\right)^{(1-\delta)/(2-\delta)} \\
&\quad + |\mathbb{E}H(X_{n,i}, X_{n,j})| \int_{\mathbb{R}^d\times\mathbb{R}^d} |H(x,y)|\, P_{X_1}(dx)\, P_{X_1}(dy)\, \|g_n\|_\infty\, n^{-1} \\
&\leq C \left(\mathbb{E}|H(X_{n,i}, X_{n,j})|^{(2-\delta)/(1-\delta)}\right)^{(1-\delta)/(2-\delta)} \bar{\tau}_{k-j}^{\delta^2} \\
&\quad + C \int_{\mathbb{R}^d\times\mathbb{R}^d} |H(x,y)|\, P_{X_1}(dx)\, P_{X_1}(dy)\, \|g_n\|_\infty\, n^{-1}\, \bar{\tau}_{j-i}^{\delta} \\
&\quad + C \left[\int_{\mathbb{R}^d\times\mathbb{R}^d} |H(x,y)|\, P_{X_1}(dx)\, P_{X_1}(dy)\, \|g_n\|_\infty\right]^2 n^{-2}.
\end{aligned}$$

Within these calculations, the inequality

$$\begin{aligned}
&\mathbb{E}\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \left|H(X_{n,k}, y) - H(\widetilde{X}_{n,k}, y)\right|\, P_{X_{n,1}}(dy) \\
&\leq \bar{\tau}_{k-j}^{\delta} \left[\int_{\mathbb{R}^d} [f_1(x,z,y,y)]^\eta (\|x\|_1 + \|z\|_1) P_{X_{n,1}}(dx) P_{X_{n,1}}(dy) P_{X_{n,1}}(dz)\right]^{1-\delta} \\
&\leq C\, \bar{\tau}_{k-j}^{\delta}
\end{aligned}$$

is applied which is a consequence of (A6)(iii), where the function f_1 is defined as in the proof of Lemma 3.8. Similarly to the above investigation of the terms E_1 up to E_4, it can be shown that there exists a family $(\varepsilon_c)_{c\in\mathbb{R}_+}$ with $\varepsilon_c \longrightarrow_{c\to\infty} 0$ such that

$$\left(\sup_{\substack{1\leq k<n \\ n\in\mathbb{N}}} \mathbb{E}|H(X_{n,1}, X_{n,1+k})|^{(2-\delta)/(1-\delta)}\right)^{(1-\delta)/(2-\delta)} + \int_{\mathbb{R}^d\times\mathbb{R}^d} |H(x,y)| P_{X_1}(dx)\, P_{X_1}(y) \leq \varepsilon_c.$$

Hence, the relation

$$\limsup_{n\to\infty} \frac{1}{(n-1)^2} \sum_{r=1}^{n-1} \sum_{\substack{1\leq i<j\leq k<l\leq n \\ r:=l-k>j-i \\ k-j\geq j-i}} \left|\mathbb{E}H(X_{n,i}, X_{n,j})H(X_{n,k}, \widetilde{X}_{n,l}^{(r)})\right| \leq C\, \varepsilon_c \underset{c\to\infty}{\longrightarrow} 0$$

holds. In order to investigate the last summand of the r.h.s. of inequality (5.3), we introduce a vector $(\widetilde{X}'_{n,k}, \widetilde{X}'_{n,j})' \stackrel{d}{=} (X'_{n,k}, X'_{n,j})'$ that is independent of $X_{n,i}$ and such that

$\mathbb{E}\|(\widetilde{X}'_{n,k},\widetilde{X}'_{n,j})'-(X'_{n,k},X'_{n,j})'\|_1\leq\bar{\tau}_{k-i}$. This implies

$$\begin{aligned}
&\left|\mathbb{E}H(X_{n,i},X_{n,j})H(X_{n,k},\widetilde{X}^{(r)}_{n,l})\right| \\
&\leq \varepsilon_c^{(1)}\,\bar{\tau}_{k-i}^{\delta 2}+\left|\mathbb{E}\int_{\mathbb{R}^d}H(y,\widetilde{X}_{n,j})\,P_{X_{n,1}}(dy)\int_{\mathbb{R}^d}H(\widetilde{X}_{n,k},y)\,P_{X_{n,1}}(dy)\right| \\
&\leq \varepsilon_c^{(1)}\,\left[\bar{\tau}_{k-i}^{\delta 2}+n^{-1}\,\|g_n\|_\infty\,\bar{\tau}_{j-k}^{\delta 2}\right]+\int_{\mathbb{R}^d\times\mathbb{R}^d}|H(x,y)|\,P_{X_1}(dx)\,P_{X_1}(dy)\,n^{-2}\,\|g_n\|_\infty^2 \\
&\leq \varepsilon_c^{(1)}\,\left[\bar{\tau}_{k-i}^{\delta 2}+n^{-1}\,\|g_n\|_\infty\,\bar{\tau}_{j-k}^{\delta 2}+n^{-2}\,\|g_n\|_\infty^2\right],
\end{aligned}$$

where $(\varepsilon_c^{(1)})_c$ is a certain family of constants with $\varepsilon_c^{(1)}\longrightarrow_{c\to\infty}0$. Consequently, we obtain

$$\limsup_{n\to\infty}\frac{1}{(n-1)^2}\sum_{r=1}^{n-1}\sum_{\substack{1\leq i\leq k<j<l\leq n\\ r:=l-j>k-i>j-k}}\left|\mathbb{E}H(X_{n,i},X_{n,j})H(X_{n,k},\widetilde{X}^{(r)}_{n,l})\right|\leq C\,\varepsilon_c^{(1)}\underset{c\to\infty}{\longrightarrow}0,$$

which completes the first step of the proof.

<u>Step 2:</u> *Wavelet approximation of the kernel.*

Denote the counterpart for triangular schemes of the U-statistic $U_{n,c}^{(L)}$, defined in the proof of Lemma 3.5, by $U_{n,n,c}^{(L)}$. Asymptotic negligibility of $\limsup_{n\to\infty}n^2\,\mathbb{E}(U_{n,n,c}-U_{n,n,c}^{(L)})^2$ can be verified using the decomposition (5.2) with $H=h_c-h_c^{(L)}$. In complete analogy to the proofs of Lemma 3.5 (step 1) and Lemma 3.8, we obtain that

$$\lim_{L\to\infty}\limsup_{n\to\infty}\sup_{1\leq k<n}8\,\mathbb{E}|H(X_{n,1},X_{n,1+k})|^2+\frac{8}{(n-1)^2}\sum_{r=1}^{n-1}\sum_{t=1}^{4}Z_{n,r}^{(t)}=0.$$

Moreover, since the functions $H^{(L)}=h_c-h_c^{(L)}$ are uniformly bounded, one can proceed as in the previous step of the proof to show that

$$\lim_{L\to\infty}\limsup_{n\to\infty}\frac{8}{(n-1)^2}\sum_{r=1}^{n-1}\sum_{t=1}^{4}\left|\widetilde{Z}_{n,r}^{(t)}-Z_{n,r}^{(t)}\right|=0.$$

Next, we introduce the U-statistic $U_{n,n,c}^{(K,L)}$ by substituting the random variables resulting from our triangular scheme in the definition of $U_{n,c}^{(K,L)}$ in Subsection 3.3.2. Similar arguments as before lead to

$$\lim_{K\to\infty}\limsup_{n\to\infty}n^2\,\mathbb{E}\left(U_{n,n,c}^{(K,L)}-U_{n,n,c}^{(L)}\right)^2=0.$$

<u>Step 3:</u> *Asymptotics of $n\,U_{n,n,c}^{(K,L)}$.*

Invoking the notation of the proof of Lemma 3.6, we obtain the modified representation

$$\begin{aligned}
&n\,U_{n,c}^{(K,L)} \\
&=\frac{1}{n-1}\sum_{i\neq j}h_c^{(K,L)}(X_{n,i},X_{n,j}) \\
&=\frac{n}{n-1}\sum_{k,l=1}^{M(K,L)}\gamma_{k,l}^{(c)}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}q_k(X_{n,i})\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}q_l(X_{n,i})\right)-\frac{1}{n}\sum_{i=1}^{n}q_k(X_{n,i})\,q_l(X_{n,i})\right].
\end{aligned}\tag{5.4}$$

Due to the continuity and boundedness of the functions q_k, the latter quantity of the r.h.s. can be reformulated as

$$\frac{1}{n}\sum_{i=1}^{n} q_k(X_{n,i})q_l(X_{n,i}) = \frac{1}{n}\sum_{i=1}^{n}[q_k(X_{n,i})q_l(X_{n,i}) - \mathbb{E}q_k(X_{n,i})q_l(X_{n,i})] \\ + \mathbb{E}q_k(X_1)q_l(X_1) + o(1),$$

where the latter sum converges to zero by Lemma 2.3. In order to investigate the remaining part of the r.h.s. of equation (5.4), we prove asymptotic normality of $n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{M(K,L)} t_k[q_k(X_{n,i}) - \mathbb{E}q_k(X_{n,i})]$, $t_1, \ldots, t_{M(K,L)} \in \mathbb{R}$. To this end, Lemma 2.2 is applied to $(Q_{n,i})_{i=1}^{n}$, $n \in \mathbb{N}$, with $Q_{n,i} = \sum_{k=1}^{M(K,L)} t_k[q_k(X_{n,i}) - \mathbb{E}q_k(X_{n,i})]$. According to boundedness and Lipschitz continuity of the functions q_k and due to the dependence structure of $(X_{n,i})_{i=1}^{n}$, $n \in \mathbb{N}$, all prerequisites of Lemma 2.2 are fulfilled if additionally

$$\left|\frac{1}{n}\operatorname{var}(Q_{n,1} + \cdots + Q_{n,n}) - \mathbb{E}Q_1Q_1 - 2\sum_{k=2}^{\infty}\operatorname{cov}(Q_1, Q_k)\right| \underset{\mathrm{n}\to\infty}{\longrightarrow} 0,$$

where the sequence $(Q_k)_{k\in\mathbb{N}}$ is defined as in the proof of Lemma 3.6. The calculations can be carried out similarly as in the proof of Theorem 4.1 and are therefore omitted. Moreover, we have $n^{-1/2}\sum_{i=1}^{n}\mathbb{E}q_k(X_{n,i}) \longrightarrow_{n\to\infty} \int_{\mathbb{R}^d} q_k(x)g(x)P_{X_1}(dx)$, $k = 1, \ldots, M(K,L)$. Thus, the continuous mapping theorem and Slutsky's theorem finally yield $n\,U_{n,n,c}^{(K,L)} \stackrel{d}{\longrightarrow} Z_{loc,c}^{(K,L)}$ with

$$\begin{aligned} Z_{loc,c}^{(K,L)} := &\sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{k_1,k_2}^{(c)}[(Z_{k_1} + C_{k_1})(Z_{k_2} + C_{k_2}) - A_{k_1,k_2}] \\ &+ \sum_{j=0}^{J(L)-1}\sum_{k_1,k_2\in\{-K,\ldots,K\}^d}\sum_{e=(e_1',e_2')'\in\bar{E}} \beta_{j;k_1,k_2}^{(c,e)}\Big[\Big(Z_{j;k_1}^{(e_1)} + D_{j;k_1}^{(e_1)}\Big) \\ &\times \Big(Z_{j;k_2}^{(e_2)} + D_{j;k_2}^{(e_2)}\Big) - B_{j;k_1,k_2}^{(e)}\Big]. \end{aligned} \tag{5.5}$$

<u>Step 4:</u> *Asymptotics of $n\,U_{n,n}$ and $n\,V_{n,n}$.*

In complete accordance to the proof of Theorem 3.1, we can deduce the asymptotic distribution of $n\,U_{n,n}$ from the previous part of the proof. In order to derive the limit distribution of $n\,V_{n,n}$, it remains to analyse

$$\frac{1}{n}\sum_{i=1}^{n} h(X_{n,i}, X_{n,i}) = \frac{1}{n}\sum_{i=1}^{n}[h(X_{n,i}, X_{n,i}) - \mathbb{E}h(X_{n,i}, X_{n,i})] + \mathbb{E}h(X_1, X_1) + o(1).$$

By Lemma 2.3 we obtain that the sum on the r.h.s. converges to zero in probability which in turn leads to $n\,V_{n,n} \xrightarrow{d} Z_{loc} + \mathbb{E}h(X_1, X_1)$. □

The three subsequent results will be invoked in the following sections to verify asymptotic unbiasedness of certain hypothesis tests against local alternatives. To this end, we first reformulate the limits of $n\,U_n$ and $n\,U_{n,n}$, respectively.

Lemma 5.1. *(i) Suppose the assumptions (A1), (A2), and (A4) hold. Then,*

$$n\,U_n \xrightarrow{d} \lim_{c\to\infty}\lim_{L\to\infty}\sum_{k_1,k_2\in\mathbb{Z}^d}\alpha^{(c)}_{J(L);k_1,k_2}\,[Z_{J(L),k_1}\,Z_{J(L),k_2}-A_{J(L);k_1,k_2}],\tag{5.6}$$

where the r.h.s. converges in the L_2-sense. Here, $(Z_{J,k})_{J,k}$ are centered and jointly normal random variables with covariance structure

$$\begin{aligned}\operatorname{cov}(Z_{J,s},Z_{J,t}) &= \operatorname{cov}(\Phi_{J,s}(X_1),\,\Phi_{J,t}(X_1))\\ &\qquad+\sum_{k=2}^{\infty}\left[\operatorname{cov}(\Phi_{J,s}(X_1),\,\Phi_{J,t}(X_k))+\operatorname{cov}(\Phi_{J,s}(X_k),\,\Phi_{J,t}(X_1))\right],\end{aligned}$$

$\alpha^{(c)}_{J;k_1,k_2} := \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c(x,y)\,\Phi_{J,k_1}(x)\,\Phi_{J,k_2}(y)\,dx\,dy$, *and* $A_{J;k_1,k_2} := \operatorname{cov}(\Phi_{J,k_1}(X_1),\Phi_{J,k_2}(X_1))$, $k,k_1,k_2\in\mathbb{Z}^d$, $J\in\mathbb{N}$. *The sequence $(J(L))_L$ and the family of functions $(h_c)_c$ are defined as in Lemma 3.8.*

(ii) Under the assumptions (A5) and (A6),

$$n\,U_{n,n} \xrightarrow{d} \lim_{c\to\infty}\lim_{L\to\infty}\sum_{k_1,k_2\in\mathbb{Z}^d}\alpha^{(c)}_{J(L);k_1,k_2}[(Z_{J(L),k_1}+C_{J(L);k_1})(Z_{J(L),k_2}+C_{J(L);k_2})-A_{J(L);k_1,k_2}],\tag{5.7}$$

where the r.h.s. converges in the L_2-sense and $C_{J;k} := \int_{\mathbb{R}^d}[\Phi_{J,k}(x)-\mathbb{E}\Phi_{J,k}(X_1)]\,g(x)\,P_{X_1}(dx)$.

Proof. We only consider (i), the part (ii) can be verified in complete analogy. Instead of a multiscale approximation of the kernel, we use only one scale J that increases to infinity.

The first two steps of the proof of Lemma 3.8 yield

$$\lim_{c\to\infty}\limsup_{n\to\infty} n^2\,\mathbb{E}\,(U_n-U_{n,c})^2=0 \quad\text{and}\quad \lim_{L\to\infty}\limsup_{n\to\infty} n^2\,\mathbb{E}\left(U_{n,c}-U_{n,c}^{(L)}\right)^2=0.$$

The U-statistic $U_{n,c}^{(L)}$ is now approximated by the U-statistic $\widetilde{U}_{n,c}^{(K,L)}$ whose kernel is defined as the degenerate version of

$$\bar{h}_c^{(K,L)}(x,y) := \sum_{k_1,k_2\in\{-K,\dots,K\}^d}\alpha^{(c)}_{J(L);k_1,k_2}\Phi_{J(L),k_1}(x)\Phi_{J(L),k_2}(y).$$

Note that this function coincides with $\widetilde{h}_c^{(L)}$, defined in the proof of Lemma 3.5, for $(x,y)\in[-B(K),B(K)]^{2d}$, where the sequence $(B(K))_K$ can be chosen such that $B(K)\to\infty$ as $K\to\infty$. Moreover, the sequence $(\bar{h}_c^{(K,L)})_K$ consists of functions that are bounded and Lipschitz continuous uniformly in K. Therefore, using the same arguments as in the proofs of Lemma 3.5 and Lemma 3.8, one obtains

$$\lim_{K\to\infty}\limsup_{n\to\infty} n^2\,\mathbb{E}\left(U_{n,c}^{(L)}-\widetilde{U}_{n,c}^{(K,L)}\right)^2=0.$$

According to the central limit theorem (Lemma 2.2 in conjunction with the Cramér-Wold device) and the continuous mapping theorem, we get

$$n\,\widetilde{U}_{n,c}^{(K,L)} \stackrel{d}{\longrightarrow} \widetilde{Z}_c^{(K,L)} := \sum_{k_1,k_2\in\{-K,\dots,K\}^d} \alpha_{J(L);k_1,k_2} Z_{J(L),k_1} Z_{J(L),k_2}. \tag{5.8}$$

In view of Billingsley [16], Theorem 4.2, it remains to show that

$$\lim_{c\to\infty}\lim_{L\to\infty}\lim_{K\to\infty}\mathbb{E}\left(\widetilde{Z}_c^{(K,L)} - Z\right)^2 = 0.$$

To this end, first note that applying the central limit theorem and the continuous mapping theorem as above, we obtain $n\,\widetilde{U}_{n,c}^{(K_1,L)} - n\,\widetilde{U}_{n,c}^{(K_2,L)} \stackrel{d}{\longrightarrow} \widetilde{Z}_c^{(K_1,L)} - \widetilde{Z}_c^{(K_2,L)}$. In order to verify that $(\widetilde{Z}_c^{(K,L)})_K$ is a Cauchy sequence in L_2, we invoke Theorem 5.3 of Billingsley [16] again:

$$\mathbb{E}\left(\widetilde{Z}_c^{(K_1,L)} - \widetilde{Z}_c^{(K_2,L)}\right)^2 \le \liminf_{n\to\infty} n^2\,\mathbb{E}\left(\widetilde{U}_{n,c}^{(K_1,L)} - \widetilde{U}_{n,c}^{(K_2,L)}\right)^2 \underset{K_1,K_2\to\infty}{\longrightarrow} 0.$$

Denoting the corresponding limit by $\widetilde{Z}_c^{(L)}$, we are to show that $(\widetilde{Z}_c^{(L)})_L$ is a Cauchy sequence in L_2. One can estimate similarly to inequality (3.16),

$$\begin{aligned}\mathbb{E}\left(\widetilde{Z}_c^{(L_1)} - \widetilde{Z}_c^{(L_2)}\right)^2 &\le 4\limsup_{K\to\infty}\mathbb{E}\left(\widetilde{Z}_c^{(K,L_1)} - \widetilde{Z}_c^{(K,L_2)}\right)^2\\ &\le 16\limsup_{K\to\infty}\liminf_{n\to\infty} n^2\,\mathbb{E}\left(\widetilde{U}_{n,c}^{(K,L_1)} - \widetilde{U}_{n,c}^{(K,L_2)}\right)^2 \underset{L_1,L_2\to\infty}{\longrightarrow} 0\end{aligned}$$

since also $n\,\widetilde{U}_{n,c}^{(K,L_1)} - n\,\widetilde{U}_{n,c}^{(K,L_2)} \stackrel{d}{\longrightarrow} \widetilde{Z}_c^{(K,L_1)} - \widetilde{Z}_c^{(K,L_2)}$. If we denote the associated limit by $\widetilde{Z}_c$ and iterate this method of proof once again, we obtain $\mathbb{E}(\widetilde{Z}_{c_1} - \widetilde{Z}_{c_2})^2 \longrightarrow 0$ as $c_1, c_2 \to \infty$. This finally yields assertion (i) of the lemma. □

Based on these representations of the limits of $n\,U_n$ and $n\,U_{n,n}$, one can deduce that Z_{loc} is stochastically larger than Z.

Lemma 5.2. *Suppose that the conditions (A2), (A5), and (A6) are satisfied. Moreover, let* $\operatorname{var}(Z) > 0$, *where the random variable* Z *is defined as in Theorem 3.1. The symmetric matrices* $\Delta_c^{(K,L)} := [\alpha^{(c)}_{J(L),k_1,k_2}]_{k_1,k_2=-K1_d}^{K1_d}$ *are assumed to be positive semidefinite for all* $c \ge c_0$, $L \ge L_0(c)$, $K \ge K_0(L,c)$ *and some* $c_0 \in \mathbb{R}_+$, $L_0, K_0 \in \mathbb{N}$. *Then,*

$$P(Z_{loc} > x) \ge P(Z > x), \quad \forall\, x \in \mathbb{R}.$$

Proof. We show that for all $\varepsilon > 0$ and $x \in \mathbb{R}$ the inequality $P(Z_{loc} > x) - P(Z > x) \ge -\varepsilon$ holds true. First note that w.l.o.g. x can assumed to be a continuity point of the cumulative distribution function of Z_{loc} since otherwise one splits up

$$P(Z_{loc} > x) - P(Z > x) \ge P(Z_{loc} > x+\delta) - P(Z > x+\delta) + P(Z > x+\delta) - P(Z > x).$$

Here, $\delta > 0$ can be chosen such that the distribution function of Z_{loc} is continuous in $x+\delta$ and $P(Z > x+\delta) - P(Z > x) \ge -\varepsilon/2$ due to the continuity of the distribution function of Z. From now on assume the distribution function of Z_{loc} to be continuous in x. According to the considerations in the proof of Lemma 5.1 and uniqueness of the weak limit, it suffices to verify

$$P\left(\widetilde{Z}_{loc,c}^{(K,L)} > x\right) - P\left(\widetilde{Z}_{c}^{(K,L)} > x\right) \geq 0, \quad c \geq c_0,\ K \geq K_0(c),\ L \geq L_0(c,K),$$

where $\widetilde{Z}_{loc,c}^{(K,L)} := \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{J;k_1,k_2}^{(c)}[(Z_{J,k_1} + C_{J;k_1})(Z_{J,k_2} + C_{J;k_2}) - A_{J;k_1,k_2}]$. Let the constants K, L, and c be fixed. Then, with $y = x + B_c^{(K,L)}$ and $B_c^{(K,L)} := \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{J;k_1,k_2}^{(c)} A_{J;k_1,k_2}$, we are to prove

$$\begin{aligned} &P\left(\left[Z^{(K,L)} + C^{(K,L)}\right]' \Delta_c^{(K,L)} \left[Z^{(K,L)} + C^{(K,L)}\right] > y\right) \\ &\geq P\left(\left[Z^{(K,L)}\right]' \Delta_c^{(K,L)} Z^{(K,L)} > y\right), \end{aligned} \tag{5.9}$$

where $Z^{(K,L)} := (Z_{J(L),-K1_d}, \ldots, Z_{J(L),K1_d})'$ and $C^{(K,L)} := (C_{J(L);-K1_d}, \ldots, C_{J(L);K1_d})'$. This is equivalent to

$$\begin{aligned} &P\left(\left[S_{(K,L)}^{1/2} Y + O'_{(K,L)} C^{(K,L)}\right]' O'_{(K,L)}\, \Delta_c^{(K,L)}\, O_{(K,L)} \left[S_{(K,L)}^{1/2} Y + O'_{(K,L)} C^{(K,L)}\right] > y\right) \\ &\geq P\left(\left[S_{(K,L)}^{1/2} Y\right]' O'_{(K,L)}\, \Delta_c^{(K,L)}\, O_{(K,L)} \left[S_{(K,L)}^{1/2} Y\right] > y\right) \end{aligned}$$

for some a diagonal matrix $S_{(K,L)}^{1/2}$ with decreasing diagonal elements $\sigma_1 \geq \cdots \geq \sigma_N > \sigma_{N+1} = \cdots = \sigma_{(2K+1)^d} = 0$ with an $N \leq (2K+1)^d$, an orthogonal matrix $O_{(K,L)}$, and a vector Y of independent standard normal variables. Note that $N \geq 1$ if the indices c, K, L are chosen sufficiently large, cf. the proof of Lemma 4.2. Introducing vectors $\bar{C}_1^{(K,L)}$ and $\bar{C}_2^{(K,L)}$ by

$$\bar{C}_{1,i}^{(K,L)} := \begin{cases} \sigma_i^{-1}\left(O'_{(K,L)} C^{(K,L)}\right)_i & \text{for } i \leq rk\left(S_{(K,L)}^{1/2}\right), \\ 0 & \text{for } i > rk\left(S_{(K,L)}^{1/2}\right), \end{cases} \qquad i = 1, \ldots, (2K+1)^d,$$

as well as

$$\bar{C}_{2,i}^{(K,L)} := \begin{cases} 0 & \text{for } i \leq rk\left(S_{(K,L)}^{1/2}\right), \\ \left(O'_{(K,L)} C^{(K,L)}\right)_i & \text{for } i > rk\left(S_{(K,L)}^{1/2}\right), \end{cases} \qquad i = 1, \ldots, (2K+1)^d,$$

we have

$$\begin{aligned} &P\left(\left[S_{(K,L)}^{1/2} Y + O'_{(K,L)} C^{(K,L)}\right]' O'_{(K,L)}\, \Delta_c^{(K,L)}\, O_{(K,L)} \left[S_{(K,L)}^{1/2} Y + O'_{(K,L)} C^{(K,L)}\right] > y\right) \\ &= P\big(\left[Y + \bar{C}_1^{(K,L)}\right]' \left[S_{(K,L)}^{1/2}\right]' O'_{(K,L)}\, \Delta_c^{(K,L)}\, O_{(K,L)} S_{(K,L)}^{1/2} \left[Y + \bar{C}_1^{(K,L)}\right] \\ &\quad + 2\left[\bar{C}_2^{(K,L)}\right]' O'_{(K,L)}\, \Delta_c^{(K,L)}\, O_{(K,L)} S_{(K,L)}^{1/2} \left[Y + \bar{C}_1^{(K,L)}\right] \\ &\quad + \left[\bar{C}_2^{(K,L)}\right]' O'_{(K,L)}\, \Delta_c^{(K,L)}\, O_{(K,L)} \bar{C}_2^{(K,L)} > y\big). \end{aligned}$$

Taking into the account the positive semidefiniteness of $\Delta_c^{(K,L)}$, the latter term can be expressed as

$$P\left(\sum_{k=1}^{(2K+1)^d} \lambda_k (Y_k+\delta_k)^2 + \sum_{k=1}^{(2K+1)^d} \eta_k > y\right) \tag{5.10}$$

with certain nonnegative constants λ_k, real-valued constants δ_k, η_k, $k=1,\ldots,(2K+1)^d$, and $\sum_{k=1}^{(2K+1)^d} \eta_k \geq 0$ while similar considerations for $\widetilde{Z}_c^{(K,L)} - B^{(K,L)}$ yield

$$P\left([Z^{(K,L)}]'\Delta_c^{(K,L)} Z^{(K,L)} > y\right) = P\left(\sum_{k=1}^{(2K+1)^d} \lambda_k Y_k^2 > y\right). \tag{5.11}$$

Note that for a standard normal random variable Y, the variable $(Y+a)^2$ is stochastically larger than Y^2 for any $a \neq 0$. Finally, since $Y_1,\ldots,Y_{(2K+1)^d}$ are independent standard normal variables, the comparison of (5.10) and (5.11) leads to inequality (5.9). □

Combining Proposition 5.1 and Lemma 5.2 yields the following assertion.

Lemma 5.3. *Under the assumptions of Lemma 5.2, the following inequality holds true:*

$$\liminf_{n\to\infty} \left[P(n\,U_{n,n} > x) - P(n\,U_n > x)\right] \geq 0, \quad \forall x \in \mathbb{R}.$$

If additionally $\mathbb{E}|h(X_1,X_1)| < \infty$, *then*

$$\liminf_{n\to\infty} P\left[(n\,V_{n,n} > x) - P(n\,V_n > x)\right] \geq 0, \quad \forall x \in \mathbb{R}.$$

Proof. Both inequalities can be verified in the same manner. Therefore, only the first proof is stated. We show that for each $\varepsilon > 0$

$$\liminf_{n\to\infty} \left[P(n\,U_{n,n} > x) - P(n\,U_n > x)\right] \geq -\varepsilon, \quad \forall x \in \mathbb{R}.$$

To this end, let the cumulative distribution function of Z_{loc} be continuous in $x+\delta$ for some $\delta > 0$ that is specified below. We obtain

$$\begin{aligned}\liminf_{n\to\infty} \left[P(n\,U_{n,n} > x) - P(n\,U_n > x)\right] &\geq \lim_{n\to\infty} \left[P(n\,U_{n,n} > x+\delta) - P(n\,U_n > x)\right] \\ &= \left[P(Z_{loc} > x+\delta) - P(Z > x)\right] \\ &\geq \left[P(Z > x+\delta) - P(Z > x)\right],\end{aligned}$$

where the latter difference is greater than $-\varepsilon$ for sufficiently small δ. □

5.2.2 Auxiliary results on U- and V-statistics with varying kernels

When the behaviour of model specification tests under local alternatives is investigated in Section 5.5, U-statistics with varying kernels occur. These kernels satisfy the following condition:

(A7) The kernels $h_n : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ are functions of the form

$$h_n(x,y) = h^{(1)}(x,y) + n^{-1/2} h^{(2)}(x,y) + n^{-1/2} h^{(2)}(y,x) + n^{-1} h^{(3)}(x,y),$$

where $h^{(1)}, h^{(2)}$, and $h^{(3)}$ satisfy the conditions (A2)(ii) and (A4). Moreover, the functions $h^{(1)}$ and $h^{(3)}$ are symmetric, $\int_{\mathbb{R}^d} h^{(1)}(x,y) P_X(dx) = 0$ and $\int_{\mathbb{R}^d} h^{(2)}(x,y) P_X(dx) = 0, \quad \forall\, y \in \mathbb{R}^d$.

We define the corresponding U- and V-statistics by

$$\bar{U}_{n,n} := \frac{1}{n(n-1)} \sum_{\substack{j,k=1 \\ j\neq k}}^{n} h_n(X_j, X_k) \quad \text{and} \quad \bar{V}_{n,n} := \frac{1}{n^2} \sum_{j,k=1}^{n} h_n(X_j, X_k).$$

Note that we do neither assume $\int_{\mathbb{R}^d} h^{(2)}(x,y)\, P_X(dy) = 0$ nor $h^{(3)}$ to be degenerate. In the special case of degenerate functions $h^{(2)}$ and $h^{(3)}$, the limit distributions of $n\,\bar{U}_{n,n}$ and $n\,\bar{V}_{n,n}$ coincide with those derived in Chapter 3 by virtue of Lemma 3.7, where $h = h^{(1)}$. In order to state a general result, some notation is introduced first. The wavelet coefficients for certain truncated versions $h_c^{(1)}$, $h_c^{(2)}$ and $h_c^{(3)}$ of the three kernels, that are specified below, are given by

$$\alpha_{k_1,k_2}^{(s,c)} := \iint_{\mathbb{R}^d \times \mathbb{R}^d} h_c^{(s)}(x,y)\, \Phi_{0,k_1}(x)\, \Phi_{0,k_2}(x)\, dx\, dy$$

and

$$\beta_{j;k_1,k_2}^{(s,c,e)} := \iint_{\mathbb{R}^d \times \mathbb{R}^d} h_c^{(s)}(x,y)\, \Psi_{j;k_1,k_2}^{(e)}(x,y)\, dx\, dy,$$

for $s = 1,2,3,\ j \in \mathbb{Z},\ k_1, k_2 \in \mathbb{Z}^d, e \in \bar{E}$.

Proposition 5.2. *Suppose that the assumptions (A1) and (A7) are fulfilled. Then,*

$$n\,\bar{U}_{n,n} \xrightarrow{d} \bar{Z}_{loc},$$

as $n \to \infty$, where

$$\begin{aligned}
\bar{Z}_{loc} := Z + \lim_{c\to\infty} \Big\{ &\sum_{k_1,k_2 \in \mathbb{Z}^d} \alpha_{k_1,k_2}^{(2,c)} \left[Z_{k_1}\, \mathbb{E}\Phi_{0,k_2}(X_1) + Z_{k_2}\, \mathbb{E}\Phi_{0,k_1}(X_1) \right] \\
&+ \sum_{j=0}^{\infty} \sum_{k_1,k_2 \in \mathbb{Z}^d} \sum_{e=(e_1',e_2')' \in \bar{E}} \beta_{j;k_1,k_2}^{(2,c,e)} \left[Z_{j;k_1}^{(e_1)}\, \mathbb{E}\Psi_{j,k_2}^{(e_2)}(X_1) + Z_{j;k_2}^{(e_2)}\, \mathbb{E}\Psi_{j,k_1}^{(e_1)}(X_1) \right] \Big\} \\
+ \lim_{c\to\infty} \Big\{ &\sum_{k_1,k_2 \in \mathbb{Z}^d} \alpha_{k_1,k_2}^{(3,c)}\, \mathbb{E}\Phi_{0,k_1}(X_1)\, \mathbb{E}\Phi_{0,k_2}(X_1) \\
&+ \sum_{j=0}^{\infty} \sum_{k_1,k_2 \in \mathbb{Z}^d} \sum_{e=(e_1',e_2')' \in \bar{E}} \beta_{j;k_1,k_2}^{(3,c,e)}\, \mathbb{E}\Psi_{j,k_1}^{(e_1)}(X_1)\, \mathbb{E}\Psi_{j,k_2}^{(e_2)}(X_1) \Big\}
\end{aligned}$$

and Z as well as $(Z_k)_k$ and $(Z_{j;k}^{(e_i)})_{j,k,e_i}$, $i = 1,2$, are defined as in Theorem 3.1 with $h = h^{(1)}$.
If additionally $\mathbb{E}|h^{(1)}(X_1,X_1)| + \mathbb{E}|h^{(2)}(X_1,X_1)| + \mathbb{E}|h^{(3)}(X_1,X_1)| < \infty$, then

$$n\,\bar{V}_{n,n} \xrightarrow{d} \bar{Z}_{loc} + \mathbb{E}h^{(1)}(X_1,X_1).$$

Proof. We proceed analogously to the proof of Proposition 5.1.

Step 1: Approximation by statistics with bounded kernels.

In order to construct a family of approximating statistics $n\bar{U}_{n,n,c}$ with bounded kernel functions $h_{n,c}$, we truncate the kernels $h^{(1)}$, $h^{(2)}$ and $h^{(3)}$ as in Lemma 3.1. That is,

$$\widetilde{h}_c^{(s)}(x,y) := \begin{cases} h^{(s)}(x,y) & \text{for } |h^{(s)}(x,y)| \le c_h^{(s)}, \\ -c_h^{(s)} & \text{for } h^{(s)}(x,y) < -c_h^{(s)}, \\ c_h^{(s)} & \text{for } h^{(s)}(x,y) > c_h^{(s)} \end{cases}$$

with $c_h^{(s)} := \max_{x,y\in[-c,c]^d} |h^{(s)}(x,y)|$ and $s = 1,2,3$. Based on these definitions, we introduce

$$\begin{aligned} h_c^{(1)}(x,y) &:= \widetilde{h}_c^{(1)}(x,y) - \int_{\mathbb{R}^d} \widetilde{h}_c^{(1)}(x,y)P_X(dy) - \int_{\mathbb{R}^d} \widetilde{h}_c^{(1)}(x,y)P_X(dx) \\ &\quad + \iint_{\mathbb{R}^d\times\mathbb{R}^d} \widetilde{h}_c^{(1)}(x,y)P_X(dx)P_X(dy), \\ h_c^{(2)}(x,y) &:= \widetilde{h}_c^{(2)}(x,y) - \int_{\mathbb{R}^d} \widetilde{h}_c^{(2)}(x,y)P_X(dx), \\ h_c^{(3)}(x,y) &:= \widetilde{h}_c^{(3)}(x,y), \end{aligned}$$

and

$$h_{n,c}(x,y) := h_c^{(1)}(x,y) + \frac{1}{\sqrt{n}}\, h_c^{(2)}(x,y) + \frac{1}{\sqrt{n}}\, h_c^{(2)}(y,x) + \frac{1}{n}\, h_c^{(3)}(x,y).$$

The U-statistics with kernels $h^{(s)}$ and $h_c^{(s)}$ are denoted by $\bar{U}_{n,n}^{(s)}$ and $\bar{U}_{n,n,c}^{(s)}$, $s = 1,2,3$, respectively. In order to show that

$$\limsup_{n\to\infty} n^2\, \mathbb{E}\left(\bar{U}_{n,n} - \bar{U}_{n,n,c}\right)^2 \underset{c\to\infty}{\longrightarrow} 0,$$

we decompose the approximation error as follows:

$$\begin{aligned} &4\sum_{s=1}^{3} n^2\, \mathbb{E}\left(\bar{U}_{n,n}^{(s)} - \bar{U}_{n,n,c}^{(s)}\right)^2 \\ &\le \sum_{s=1}^{3} \frac{C}{(n-1)^{1+s}}\, \mathbb{E}\left(\sum_{\substack{j,k=1\\ j\ne k}}^{n} h^{(s)}(X_j,X_k) - h_c^{(s)}(X_j,X_k)\right)^2. \end{aligned} \tag{5.12}$$

The calculations for $s = 1$ are identical with those of Lemma 3.1 and Lemma 3.8. Therefore, they are omitted. Using the abbreviation $H_c^{(2)} := h^{(2)} - h_c^{(2)}$, the second summand on the r.h.s. of inequality (5.12) can be estimated from above by

$$\begin{aligned}
&\frac{C}{(n-1)^3} \sum_{\substack{i,j,k,l=1\\ i\neq j;k\neq l}}^{n} \left|\mathbb{E}H^{(2)}(X_i,X_j)H^{(2)}(X_k,X_l)\right| \\
&\leq \frac{C}{(n-1)^3} \sum_{\substack{i,j,k,l=1\\ i<j,k,l\\ k\neq l}}^{n} \left|\mathbb{E}H^{(2)}(X_i,X_j)H^{(2)}(X_k,X_l)\right| \\
&\quad + \frac{C}{(n-1)^3} \sum_{\substack{i,j,k,l=1\\ i>j,k,l\\ k\neq l}}^{n} \left|\mathbb{E}H^{(2)}(X_i,X_j)H^{(2)}(X_k,X_l)\right| \\
&\quad + \frac{C}{(n-1)^3} \sum_{\substack{i,j,k,l=1\\ j<i<k<l}}^{n} \left|\mathbb{E}H^{(2)}(X_i,X_j)H^{(2)}(X_k,X_l)\right| \\
&\quad + \frac{C}{(n-1)^3} \sum_{\substack{i,j,k,l=1\\ j<k<i<l}}^{n} \left|\mathbb{E}H^{(2)}(X_i,X_j)H^{(2)}(X_k,X_l)\right| + \varepsilon_c \\
&=: E_1 + E_2 + E_3 + E_4 + \varepsilon_c
\end{aligned}$$

for some family $(\varepsilon_c)_c$ with $\varepsilon_c \longrightarrow_{c\to\infty} 0$. The first two terms, E_1 and E_2, can be treated similarly. For that reason only E_2 is taken into further consideration. To this end, we introduce random variables $\widetilde{X}_i \sim P_X$ that are independent of $(X_j', X_k', X_l')'$ and such that $\mathbb{E}\|X_i - \widetilde{X}_i\|_1 \leq \tau_{i-\max\{j,k,l\}}$. An iterative application of Hölder's inequality yields

$$\begin{aligned}
E_2 &= \frac{C}{(n-1)^3} \sum_{\substack{i,j,k,l=1\\ i>j,k,l\\ k\neq l}}^{n} \left|\mathbb{E}H^{(2)}(X_k,X_l)[H^{(2)}(X_i,X_j) - H^{(2)}(\widetilde{X}_i,X_j)]\right| \\
&\leq C \sum_{r=1}^{n} \tau_r^{\delta^2} \left(\sup_{k\in\mathbb{N}} \mathbb{E}\left|H^{(2)}(X_1,X_{1+k})\right|^{(2-\delta)/(1-\delta)}\right)^{(1-\delta)/(2-\delta)} \\
&\leq C \sum_{r=1}^{n} \tau_r^{\delta^2} \left(\sup_{k\in\mathbb{N}} \mathbb{E}\left|h^{(2)}(X_1,X_{1+k})\right|^{\nu} + \mathbb{E}\left|h^{(2)}(X_1,\widetilde{X}_1)\right|^{\nu}\right)^{1/\nu} \\
&\qquad \times \left(P(X_1 \notin [-c,c]^d)\right)^{[\nu(1-\delta)-(2-\delta)]/[\nu(2-\delta)]} \\
&\leq \widetilde{\varepsilon}_c
\end{aligned}$$

for some family $(\widetilde{\varepsilon}_c)_c$ with $\widetilde{\varepsilon}_c \longrightarrow_{c\to\infty} 0$. In order to derive a similar result for E_3, we introduce random vectors $(\widetilde{X}_k', \widetilde{X}_l')' \sim P_{X_k,X_l}$ that are independent of $(X_j', X_i')'$ and such that $\mathbb{E}\|\widetilde{X}_k - X_k\|_1 + \mathbb{E}\|\widetilde{X}_l - X_l\|_1 \leq \tau_{k-i}$. Moreover, let $\widetilde{X}_i \sim P_X$ be independent of X_j and such that $\mathbb{E}\|X_i - \widetilde{X}_i\|_1 \leq \tau_{i-j}$. With similar arguments as before, the expression E_3

can be estimated as follows:

$$
\begin{aligned}
E_3 &\leq \frac{C}{(n-1)^3} \sum_{1\leq j<i<k<l\leq n} \mathbb{E}\left|H^{(2)}(X_i,X_j)\right| \left|H^{(2)}(X_k,X_l) - H^{(2)}(\widetilde{X}_k,\widetilde{X}_l)\right| \\
&\quad + \frac{C}{(n-1)^3} \sum_{1\leq j<i<k<l\leq n} \left|\mathbb{E}H^{(2)}(X_i,X_j)\right| \left|\mathbb{E}H^{(2)}(X_k,X_l)\right| \\
&\leq C\,\widetilde{\varepsilon}_c + \frac{C\,\widetilde{\varepsilon}_c}{(n-1)} \sum_{1\leq j<i\leq n}^{n} \left|\mathbb{E}\big[H^{(2)}(X_i,X_j) - H^{(2)}(\widetilde{X}_i,X_j)\big]\right| \\
&\leq C\,\widetilde{\varepsilon}_c.
\end{aligned}
$$

It remains to prove a similar inequality for E_4. To this end, the introduction of the following random variables is useful: First, let $\widetilde{X}_l \sim P_X$ be independent of $(X_j', X_k', X_i')'$ and such that $\mathbb{E}\|X_l - \widetilde{X}_l\|_1 \leq \tau_{l-i}$, then introduce $\widetilde{X}_i \sim P_X$ independently of $(X_j', X_k')'$ and with $\mathbb{E}\|X_i - \widetilde{X}_i\|_1 \leq \tau_{i-k}$. In accordance with the foregoing investigations, the quantity E_4 can be bounded from above by

$$
\begin{aligned}
E_4 &\leq C\,\widetilde{\varepsilon}_c + \frac{C}{(n-1)^3} \sum_{1\leq j<k<i<l\leq n} \left|\mathbb{E}H^{(2)}(X_i,X_j)H^{(2)}(X_k,\widetilde{X}_l)\right| \\
&\leq C\,\widetilde{\varepsilon}_c + \frac{C}{(n-1)^2} \sum_{1\leq j<k<i\leq n} \left|\mathbb{E}\big[H^{(2)}(X_i,X_j) - H^{(2)}(\widetilde{X}_i,X_j)\big] \int_{\mathbb{R}^d} H^{(2)}(X_k,z)P_X(dz)\right| \\
&\leq C\,\widetilde{\varepsilon}_c.
\end{aligned}
$$

Summing up, the second summand of inequality (5.12) decreases to zero with increasing truncation parameter c uniformly in $n \in \mathbb{N}$. The consideration of its third summand is rather easy. Denoting an independent copy of X_1 by $\widetilde{X}_1$, we obtain

$$
\begin{aligned}
&\frac{C}{(n-1)^4}\,\mathbb{E}\left(\sum_{\substack{j,k=1\\ j\neq k}}^{n} h^{(3)}(X_j,X_k) - h_c^{(3)}(X_j,X_k)\right)^2 \\
&\leq C\left(\sup_{k\in\mathbb{N}} \mathbb{E}\left[h^{(3)}(X_1,X_{1+k}) - h_c^{(3)}(X_1,X_{1+k})\right]^2 + \mathbb{E}\left[h^{(3)}(X_1,\widetilde{X}_1) - h_c^{(3)}(X_1,\widetilde{X}_1)\right]^2\right),
\end{aligned}
$$

which vanishes asymptotically with increasing truncation parameter c uniformly in n.
Step 2: *Wavelet approximation of the kernel.*
Using the same decompositions and couplings as in previous step as well as the estimation arguments of the proofs of Lemma 3.5 and Lemma 3.8, we get $\lim_{L\to\infty}\limsup_{n\to\infty}\sum_{s=1}^{3} n^2\,\mathbb{E}(\bar{U}_{n,n,c}^{(s,L)} - \bar{U}_{n,n,c}^{(s)})^2 = 0$ and $\lim_{K\to\infty}\limsup_{n\to\infty}\sum_{s=1}^{3} n^2\,\mathbb{E}(\bar{U}_{n,n,c}^{(s,K,L)} - \bar{U}_{n,n,c}^{(L,s)})^2 = 0$. Here, the statistics $\bar{U}_{n,n,c}^{(L,s)}$ are constructed from $\widetilde{h}_{c,L}^{(s)}$ in the same way as $\bar{U}_{n,n,c}^{(s)}$ have been built on the basis of $\widetilde{h}_c^{(s)}$, $s = 1,2,3$, where $\widetilde{h}_{c,L}^{(s)} := \sum_{k_1,k_2\in\mathbb{Z}^d}\alpha_{J(L);k_1,k_2}^{(s,c)}\Phi_{J(L),k_1}\Phi_{J(L),k_2}$ and

$\alpha^{(s,c)}_{J(L);k_1,k_2} := \int_{\mathbb{R}^d\times\mathbb{R}^d} h^{(s)}_c(x,y)\,\Phi_{J(L),k_1}(x)\,\Phi_{J(L),k_2}(y)\,dx\,dy$. Moreover,

$$
\begin{aligned}
(n-1)\,\bar U^{(1,K,L)}_{n,n,c} := \sum_{k_1,k_2\in\{-K,\dots,K\}^d} & \alpha^{(1,c)}_{k_1,k_2}\Big\{\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n[\Phi_{0,k_1}(X_i)-\mathbb{E}\Phi_{0,k_1}(X_1)]\Big)\\
&\times\quad\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n[\Phi_{0,k_2}(X_i)-\mathbb{E}\Phi_{0,k_2}(X_1)]\Big)\\
&-\frac{1}{n}\sum_{i=1}^n[\Phi_{0,k_1}(X_i)-\mathbb{E}\Phi_{0,k_1}(X_1)]\,[\Phi_{0,k_2}(X_i)-\mathbb{E}\Phi_{0,k_2}(X_1)]\Big\}\\
+\sum_{j=0}^{J(L)-1}\sum_{k_1,k_2\in\{-K,\dots,K\}^d}\sum_{e\in\bar E} & \beta^{(1,c,e)}_{j;k_1,k_2}\Big\{\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n\big[\Psi^{(e_1)}_{j,k_1}(X_i)\\
&-\mathbb{E}\Psi^{(e_1)}_{j,k_1}(X_1)\big]\Big)\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n\big[\Psi^{(e_2)}_{j,k_2}(X_i)-\mathbb{E}\Psi^{(e_2)}_{j,k_2}(X_1)\big]\Big)\\
&-\frac{1}{n}\sum_{i=1}^n\big[\Psi^{(e_1)}_{j,k_1}(X_i)-\mathbb{E}\Psi^{(e_1)}_{j,k_1}(X_1)\big]\big[\Psi^{(e_2)}_{j,k_2}(X_i)-\mathbb{E}\Psi^{(e_2)}_{j,k_2}(X_1)\big]\Big\},
\end{aligned}
$$

$$
\begin{aligned}
(n-1)\,\bar U^{(2,K,L)}_{n,n,c} := \sum_{k_1,k_2\in\{-K,\dots,K\}^d} & \alpha^{(2,c)}_{k_1,k_2}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n[\Phi_{0,k_1}(X_i)-\mathbb{E}\Phi_{0,k_1}(X_1)]\\
&\times\frac{1}{n}\sum_{\substack{k=1\\k\neq i}}^n\Phi_{0,k_2}(X_k)\Big)\\
+\sum_{j=0}^{J(L)-1}\sum_{k_1,k_2\in\{-K,\dots,K\}^d}\sum_{e\in\bar E} & \beta^{(2,c,e)}_{j;k_1,k_2}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n\big[\Psi^{(e_1)}_{j,k_1}(X_i)-\mathbb{E}\Psi^{(e_1)}_{j,k_1}(X_1)\big]\\
&\times\Big[\frac{1}{n}\sum_{\substack{k=1\\k\neq i}}^n\Psi^{(e_2)}_{j,k_2}(X_k)\Big]\Big),
\end{aligned}
$$

and

$$
\begin{aligned}
&(n-1)\,\bar U^{(3,K,L)}_{n,n,c}\\
&:=\sum_{k_1,k_2\in\{-K,\dots,K\}^d}\alpha^{(3,c)}_{k_1,k_2}\left(\frac{1}{n}\sum_{i=1}^n\Phi_{0,k_1}(X_i)\left(\frac{1}{n}\sum_{\substack{k=1\\k\neq i}}^n\Phi_{0,k_2}(X_k)\right)\right)\\
&\quad+\sum_{j=0}^{J(L)-1}\sum_{k_1,k_2\in\{-K,\dots,K\}^d}\sum_{e\in\bar E}\beta^{(3,c,e)}_{j;k_1,k_2}\left(\frac{1}{n}\sum_{i=1}^n\Psi^{(e_1)}_{j,k_1}(X_i)\left(\frac{1}{n}\sum_{\substack{k=1\\k\neq i}}^n\Psi^{(e_2)}_{j,k_2}(X_k)\right)\right).
\end{aligned}
$$

Step 3: Asymptotics of $n\,\bar U_{n,n}$ and $n\,\bar V_{n,n}$.
Applying a central limit theorem (Lemma 2.2), the continuous mapping theorem, and the law of large numbers (Lemma 2.3) to $n\,\bar U^{(K,L)}_{n,n,c} = n\,[\bar U^{(1,K,L)}_{n,n,c} + 2\,\bar U^{(2,K,L)}_{n,n,c} + \bar U^{(3,K,L)}_{n,n,c}]$, we

obtain $n\,\bar{U}_{n,n,c}^{(K,L)} \xrightarrow{d} \bar{Z}_{loc,c}^{(K,L)}$, where

$$\begin{aligned}
\bar{Z}_{loc,c}^{(K,L)} := & \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{k_1,k_2}^{(1,c)} \left[Z_{k_1}\, Z_{k_2} - A_{k_1,k_2}\right] \\
& + \sum_{j=0}^{J(L)-1} \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \sum_{e\in\bar{E}} \beta_{j;k_1,k_2}^{(1,c,e)} \left[Z_{j;k_1}^{(e_1)}\, Z_{j;k_2}^{(e_2)} - B_{j;k_1,k_2}\right] \\
& + \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{k_1,k_2}^{(2,c)} \left[Z_{k_1}\,\mathbb{E}\Phi_{0,k_2}(X_1) + Z_{k_2}\,\mathbb{E}\Phi_{0,k_1}(X_1)\right] \\
& + \sum_{j=0}^{J(L)-1} \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \sum_{e\in\bar{E}} \beta_{j;k_1,k_2}^{(2,c,e)} \left[Z_{j;k_1}^{(e_1)}\,\mathbb{E}\Psi_{j,k_2}^{(e_2)}(X_1) + Z_{j;k_2}^{(e_2)}\,\mathbb{E}\Psi_{j,k_1}^{(e_1)}(X_1)\right] \\
& + \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{k_1,k_2}^{(3,c)}\,\mathbb{E}\Phi_{0,k_1}(X_1)\,\mathbb{E}\Phi_{0,k_2}(X_1) \\
& + \sum_{j=0}^{J(L)-1} \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \sum_{e\in\bar{E}} \beta_{j;k_1,k_2}^{(3,c,e)}\,\mathbb{E}\Psi_{j,k_1}^{(e_1)}(X_1)\,\mathbb{E}\Psi_{j,k_2}^{(e_2)}(X_1).
\end{aligned} \tag{5.13}$$

Thus, standard arguments lead to the desired limit of $n\,\bar{U}_{n,n}$. Moreover, the law of large numbers and Slutsky's theorem imply

$$\begin{aligned}
n\bar{V}_{n,n} &= (n-1)\,\bar{U}_{n,n} + \frac{1}{n}\sum_{k=1}^{n}\left[h^{(1)}(X_k,X_k) + n^{-1/2}\,h^{(2)}(X_k,X_k) + n^{-1}\,h^{(3)}(X_k,X_k)\right] \\
&\xrightarrow{d} \bar{Z}_{loc} + \mathbb{E}h^{(1)}(X_1,X_1),
\end{aligned}$$

which completes the proof. □

We intend to show that $\bar{Z}_{loc}$ is stochastically larger than Z for a special class of kernel functions. For this purpose an alternative representation of the limit of $n\,\bar{U}_{n,n}$ is established first.

Lemma 5.4. *Suppose that the assumptions (A1) and (A7) hold. Then,*

$$\begin{aligned}
\bar{Z}_{loc} \overset{d}{=} & \lim_{c\to\infty}\lim_{L\to\infty} \sum_{k_1,k_2\in\mathbb{Z}^d} \alpha_{J(L);k_1,k_2}^{(1,c)} \left[Z_{J(L),k_1} Z_{J(L),k_2} - \operatorname{cov}(\Phi_{J(L),k_1},\Phi_{J(L),k_2})\right] \\
& + \lim_{c\to\infty}\lim_{L\to\infty} \sum_{k_1,k_2\in\mathbb{Z}^d} \alpha_{J(L);k_1,k_2}^{(2,c)} \left[Z_{J(L),k_1}\,\mathbb{E}\Phi_{J(L),k_2}(X_1) + Z_{J(L),k_2}\,\mathbb{E}\Phi_{J(L),k_1}(X_1)\right] \\
& + \lim_{c\to\infty}\lim_{L\to\infty} \sum_{k_1,k_2\in\mathbb{Z}^d} \alpha_{J(L);k_1,k_2}^{(3,c)}\,\mathbb{E}\Phi_{J(L),k_1}(X_1)\,\mathbb{E}\Phi_{J(L),k_2}(X_1),
\end{aligned} \tag{5.14}$$

where $\alpha_{J;k_1,k_2}^{(s,c)} := \iint_{\mathbb{R}^d\times\mathbb{R}^d} h_c^{(s)}(x,y)\,\Phi_{J,k_1}(x)\,\Phi_{J,k_2}(y)\,dx\,dy$ and $(J(L))_L \subseteq \mathbb{N}$ is defined as in Lemma 3.8. The variables $(Z_{J,k})_{J,k}$ are defined as is Lemma 5.1

Proof. According to the uniqueness of the weak limit it suffices to show that $n\,\bar{U}_{n,n} \xrightarrow{d} \bar{\bar{Z}}_{loc}$, where $\bar{\bar{Z}}_{loc}$ abbreviates the r.h.s. of the equation (5.14). In view of the proof of the previous proposition, the relations $\lim_{c\to\infty}\limsup_{n\to\infty}\sum_{s=1}^{3} n^2\,\mathbb{E}(\bar{U}_{n,n,c}^{(s)}-\bar{U}_{n,n}^{(s)})^2=0$ and $\lim_{L\to\infty}\limsup_{n\to\infty}\sum_{s=1}^{3} n^2\,\mathbb{E}(\bar{U}_{n,n,c}^{(s,L)}-\bar{U}_{n,n,c}^{(s)})^2=0$ hold true. Now we define

$$(n-1)\,\bar{\bar{U}}_{n,n,c}^{(1,K,L)} := \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{J(L);k_1,k_2}^{(1,c)} \Big\{\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\Phi_{J(L),k_1}(X_i)-\mathbb{E}\Phi_{J(L),k_1}(X_1)\Big]\Big)$$
$$\times\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\Phi_{J(L),k_2}(X_i)-\mathbb{E}\Phi_{J(L),k_2}(X_1)]\Big)$$
$$-\frac{1}{n}\sum_{i=1}^{n}[\Phi_{J(L),k_1}(X_i)-\mathbb{E}\Phi_{J(L),k_1}(X_1)][\Phi_{J(L),k_2}(X_i)-\mathbb{E}\Phi_{J(L),k_2}(X_1)]\Big\},$$

$$(n-1)\,\bar{\bar{U}}_{n,n,c}^{(2,K,L)} := \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{J(L);k_1,k_2}^{(2,c)} \Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\Phi_{J(L),k_1}(X_i)-\mathbb{E}\Phi_{J(L),k_1}(X_1)]$$
$$\times\frac{1}{n}\sum_{\substack{k=1\\k\neq i}}^{n}\Phi_{J(L),k_2}(X_k)\Big),$$

and

$$(n-1)\,\bar{\bar{U}}_{n,n,c}^{(3,K,L)} := \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{J(L);k_1,k_2}^{(3,c)} \Big[\frac{1}{n}\sum_{i=1}^{n}\Phi_{J(L),k_1}(X_i)\Big(\frac{1}{n}\sum_{\substack{k=1\\k\neq i}}^{n}\Phi_{J(L),k_2}(X_k)\Big)\Big].$$

Invoking the same arguments as before, one can show that $\lim_{K\to\infty}\limsup_{n\to\infty}\sum_{s=1}^{3} n^2\,\mathbb{E}(\bar{\bar{U}}_{n,n,c}^{(s,K,L)}-\bar{U}_{n,n,c}^{(L,s)})^2 = 0$ since the corresponding kernels (before their artificial degeneration in the cases $s=1$ and $s=2$) coincide on any multidimensional interval $[-B,B]^{2d}$ if K is chosen sufficiently large. Moreover, the Cramér-Wold device, Lemma 2.2, Lemma 2.3, and the continuous mapping theorem imply that $n\,[\bar{\bar{U}}_{n,n,c}^{(1,K,L)}+2\bar{\bar{U}}_{n,n,c}^{(2,K,L)}+\bar{\bar{U}}_{n,n,c}^{(3,K,L)}] \xrightarrow{d} \bar{\bar{Z}}_{loc,c}^{(K,L)}$, where

$$\begin{aligned}\bar{\bar{Z}}_{loc,c}^{(K,L)} :=& \sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{J(L);k_1,k_2}^{(1,c)}\left[Z_{J(L),k_1}\,Z_{J(L),k_2}-\operatorname{cov}(\Phi_{J(L),k_1}(X_1),\Phi_{J(L),k_2}(X_1))\right]\\ &+\sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{J(L);k_1,k_2}^{(2,c)}\left[Z_{J(L),k_1}\mathbb{E}\Phi_{J(L),k_2}(X_1)+Z_{J(L),k_2}\mathbb{E}\Phi_{J(L),k_1}(X_1)\right]\\ &+\sum_{k_1,k_2\in\{-K,\ldots,K\}^d} \alpha_{J(L);k_1,k_2}^{(3,c)}\mathbb{E}\Phi_{J(L),k_1}(X_1)\,\mathbb{E}\Phi_{J(L),k_2}(X_1).\end{aligned} \tag{5.15}$$

Now we are in the position to deduce the assertion of the lemma. To this end, one merely has to follow the lines of the proof of Lemma 5.1. □

To prove that $\bar{Z}_{loc}$ is stochastically larger than Z, we assume:

(A8) **(i)** The kernels $h^{(s)}:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ satisfy (A7) and have the forms $h^{(1)}(x,y)=\sum_{i=1}^{M}\int_{\mathbb{R}^q} g_i^{(1)}(x,t)g_i^{(1)}(y,t)\,dt$, $h^{(2)}(x,y)=\sum_{i=1}^{M}\int_{\mathbb{R}^q} g_i^{(1)}(x,t)g_i^{(2)}(y,t)\,dt$, and

$h^{(3)}(x,y) = \sum_{i=1}^{M} \int_{\mathbb{R}^q} g_i^{(2)}(x,t) g_i^{(2)}(y,t)\, dt$, where $g_i^{(s_j)} : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}$ are continuous functions such that $\int_{\mathbb{R}^q} \max_{x,y\in[-c,c]^d} |g_i^{(s_1)}(x,t)\, g_i^{(s_2)}(y,t)| dt < \infty,\ \forall\, c \in \mathbb{R}_+,\ s_1, s_2 = 1,2,\ i = 1,\dots,M,\ M \in \mathbb{N}$.

(ii) For some δ satisfying (A7), a $\nu > (2-\delta)/(1-\delta)$, and $s_1, s_2 = 1,2,\ i = 1,\dots,M$,

$$\sup_{k\in\mathbb{N}} \mathbb{E}\left(\int_{\mathbb{R}^q} |g_i^{(s_1)}(X_1,t) g_i^{(s_2)}(X_{1+k},t)| dt\right)^{\nu} + \mathbb{E}\left(\int_{\mathbb{R}^q} |g_i^{(s_1)}(X_1,t) g_i^{(s_2)}(\widetilde{X}_1,t)| dt\right)^{\nu}$$

is finite. Here, $\widetilde{X}_1$ denotes an independent copy of X_1.

(iii) There exist continuous functions $f_i^{(s)} : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^q$ that are symmetric in their first two arguments such that for all $s = 1,2,\ i = 1,\dots,M$,

$$|g_i^{(s)}(x,t) - g_i^{(s)}(\bar{x},t)| \le f_i^{(s)}(x,\bar{x},t) \|x - \bar{x}\|_1.$$

Moreover,

$$\sup_{k_1,\dots,k_4\in\mathbb{N}} \mathbb{E}\Big(\max_{a_1,a_2\in[-A,A]^d} \Big[\int_{\mathbb{R}^q} f_i^{(s_1)}(Y_{k_1}, Y_{k_2} + a_1, t) \times |g_i^{(s_2)}(Y_{k_3} + a_2, t)| dt \Big]^{\eta} \|Y_{k_4}\|_1 \Big) < \infty, \quad s_1, s_2 = 1,2,$$

for $\eta := 1/(1-\delta)$, some $A > 0$ and any $(Y'_{k_1},\dots,Y'_{k_4})'$ consisting of independent subvectors $(Y'_{k_{j_1(m)}},\dots,Y'_{k_{j_l(m)}})' \stackrel{d}{=} (X'_{k_{j_1(m)}},\dots,X'_{k_{j_l(m)}})',\ l,m = 1,\dots,4$.

Lemma 5.5. *Suppose that the conditions (A1) and (A8) are satisfied. If additionally* $\operatorname{var}(Z) > 0$, *then*

$$P(\bar{Z}_{loc} > x) \ge P(Z > x), \quad \forall\, x \in \mathbb{R}.$$

Proof. According to Remark 3.3 one can use any other truncation method instead of the one proposed in Lemma 3.1 as long as the conditions 1. to 4. of Remark 3.3 are satisfied. Obviously, the same holds true for the kernel truncations in the proof of Lemma 3.8, and thus of Proposition 5.2, if the validity of (A4)(i) is preserved uniformly for the truncated kernels instead of 3. in Remark 3.3. The following method to construct bounded approximating kernels turns out to be suitable in the present context: For all $c > 0,\ t \in \mathbb{R}^q$, and $s = 1,2,\ i = 1,\dots,M$, define $c_i^{(s)}(t) := \max_{x\in[-c,c]^d} |g_i^{(s)}(x,t)|$,

$$g_{i,c}^{(s)}(x,t) := \begin{cases} g_i^{(s)}(x,t) & \text{for } |g_i^{(s)}(x,t)| \le c_i^{(s)}(t), \\ -c_i^{(s)}(t) & \text{for } g_i^{(s)}(x,t) < -c_i^{(s)}(t), \\ c_i^{(s)}(t) & \text{for } g_i^{(s)}(x,t) > c_i^{(s)}(t), \end{cases}$$

and $\bar{h}_c^{(1)}(x,y) := \sum_{i=1}^{M} \int_{\mathbb{R}^q} g_{i,c}^{(1)}(x,t)\, g_{i,c}^{(1)}(y,t)\, dt$, $\bar{h}^{(2)}(x,y) := \sum_{i=1}^{M} \int_{\mathbb{R}^q} g_{i,c}^{(1)}(x,t)\, g_{i,c}^{(2)}(y,t)\, dt$, as well as $\bar{h}_c^{(3)}(x,y) := \sum_{i=1}^{M} \int_{\mathbb{R}^q} g_{i,c}^{(2)}(x,t)\, g_{i,c}^{(2)}(y,t)\, dt$. In view of assumption (A8), the conditions 1., 2. and 4. of Remark 3.3 are satisfied and (A4)(i) is preserved uniformly in c. Plugging in these new truncated versions of the kernels into Lemma 3.8 and Proposition 5.2 instead of $\widetilde{h}_c^{(s)}$, we obtain

$$\lim_{c\to\infty}\lim_{L\to\infty}\lim_{K\to\infty}\mathbb{E}\left(\bar{Z}-\widetilde{\widetilde{Z}}_c^{(K,L)}\right)^2=0 \quad\text{and}\quad \lim_{c\to\infty}\lim_{L\to\infty}\lim_{K\to\infty}\mathbb{E}\left(\widetilde{\widetilde{Z}}_{loc,c}^{(K,L)}-\bar{\bar{Z}}_{loc}\right)^2=0,$$

where $\widetilde{\widetilde{Z}}_c^{(K,L)}$ and $\widetilde{\widetilde{Z}}_{loc,c}^{(K,L)}$ originate from the substitution of $\alpha_{J(L);k_1,k_2}^{(c)}$ and $\alpha_{J(L);k_1,k_2}^{(s,c)}$ by

$$\bar{\alpha}_{J(L);k_1,k_2}^{(c)}:=\iint_{\mathbb{R}^d\times\mathbb{R}^d}\bar{h}_c(x,y)\,\Phi_{J(L),k_1}(x)\,\Phi_{J(L),k_2}(y)\,dx\,dy$$

and

$$\bar{\alpha}_{J(L);k_1,k_2}^{(s,c)}:=\iint_{\mathbb{R}^d\times\mathbb{R}^d}\bar{h}_c^{(s)}(x,y)\,\Phi_{J(L),k_1}(x)\,\Phi_{J(L),k_2}(y)\,dx\,dy,\quad s=1,2,3,$$

in the definitions of $\widetilde{Z}_c^{(K,L)}$ and $\bar{\bar{Z}}_{loc,c}^{(K,L)}$, see the equations (5.8) and (5.15), respectively. Here, $\bar{Z}$ and $\bar{\bar{Z}}_{loc}$ abbreviate the r.h.s. of (5.6) and (5.14), respectively.

As we have already discussed in the proof of Lemma 5.2, it suffices to verify that $P(\widetilde{\widetilde{Z}}_{loc,c}^{(K,L)}>x)\geq P(\widetilde{\widetilde{Z}}_c^{(K,L)}>x)$ for sufficiently large $c>0$, $K,L\in\mathbb{N}$ and all continuity points x of the cumulative distribution function of $\bar{Z}_{loc}$. Note that the summands of $\widetilde{\widetilde{Z}}_{loc,c}^{(K,L)}$ and $\widetilde{\widetilde{Z}}_c^{(K,L)}$ that contain $\mathrm{cov}(\Phi_{J(L),k_1}(X_1),\Phi_{J(L),k_2}(X_1))$ are identical. Thus, they can be omitted in all further calculations. We denote the corresponding counterparts of $\widetilde{\widetilde{Z}}_{loc,c}^{(K,L)}$ and $\widetilde{\widetilde{Z}}_c^{(K,L)}$ by $\widehat{Z}_{loc,c}^{(K,L)}$ and $\widehat{Z}_c^{(K,L)}$, respectively. This allows for the reformulations

$$\widehat{Z}_c^{(K,L)}=\sum_{i=1}^M\int_{\mathbb{R}^q}\Big(\sum_{k\in\{-K,\dots,K\}^d}\int_{\mathbb{R}^d}Z_{J(L),k}\,g_{i,c}^{(1)}(x,t)\,\Phi_{J(L),k}(x)\,dx\Big)^2dt$$

and

$$\begin{aligned}\widehat{Z}_{loc,c}^{(K,L)}=&\sum_{i=1}^M\int_{\mathbb{R}^q}\Big(\sum_{k\in\{-K,\dots,K\}^d}\int_{\mathbb{R}^d}\Big[Z_{J(L),k}\,g_{i,c}^{(1)}(x,t)\\&+\mathbb{E}\Phi_{J(L),k}(X_1)\,g_{i,c}^{(2)}(x,t)\Big]\Phi_{J(L),k}(x)\,dx\Big)^2dt.\end{aligned}$$

Thus, there exist a vector Y consisting of i.i.d. standard normal variables $(Y_k)_{k=1}^{(2K+1)^d}$, a diagonal matrix Λ, a vector B, and a constant C such that

$$\widehat{Z}_c^{(K,L)}\overset{d}{=}Y'\Lambda Y\geq 0\quad\text{and}\quad\widehat{Z}_{loc,c}^{(K,L)}\overset{d}{=}Y'\Lambda Y+B'Y+C\geq 0\quad\text{a.s.}$$

Therefore, there are constants a_k,b_k,m with $a_k,m\geq 0$ such that

$$\widehat{Z}_c^{(K,L)}\overset{d}{=}\sum_{k=1}^{(2K+1)^d}a_kY_k^2\quad\text{and}\quad\widehat{Z}_{loc,c}^{(K,L)}\overset{d}{=}\sum_{k=1}^{(2K+1)^d}a_k\left[Y_k+b_k\right]^2+m.$$

Now the assertion follows invoking similar arguments as in the proof of Lemma 5.2. □

Lemma 5.6. *Under the assumptions of Lemma 5.5 the following inequality holds true:*

$$\liminf_{n\to\infty} \left[P(n\,\bar{U}_{n,n} > x) - P(n\,U_n > x)\right] \geq 0, \quad \forall\, x \in \mathbb{R},$$

and if additionally $\mathbb{E}|h^{(1)}(X_1, X_1)| + \mathbb{E}|h^{(2)}(X_1, X_1)| + \mathbb{E}|h^{(3)}(X_1, X_1)| < \infty$,

$$\liminf_{n\to\infty} \left[P(n\,\bar{V}_{n,n} > x) - P(n\,V_n > x)\right] \geq 0, \quad \forall\, x \in \mathbb{R}.$$

Proof. The proof is identical to the one of Lemma 5.3 and therefore omitted. □

5.3 Testing symmetry

5.3.1 Motivation

The distribution of some $\mathbb{R}^d$-valued random variable X is said to be *symmetric* about $\mu \in \mathbb{R}^d$ if $P_{X-\mu} = P_{-(X-\mu)}$. Often robust estimators of location, e.g. trimmed means, and robust tests for location parameters assume the observations to arise from a symmetric distribution, see for instance Staudte and Sheather [103]. Consequently, it is important to check this assumption before applying those methods.

Moreover, symmetry is an essential issue for numerous economic models. The validity of the CAPM (capital asset pricing model) is based on the assumption of symmetric underlying asset returns, see Lee [85] or Hodgson, Linton and Vorkink [73]. The role of symmetry in business circles was discussed by Ramsey and Rothman [98]: It can often be observed that up-wings in economic time series are longer and slower than down wings. This is contrary to a symmetry in time, which is referred to as time reversibility. By definition, a stationary time series $(X_t)_{t\in T}$ (e.g. $T = \mathbb{N}$ or $T = \mathbb{R}_0^+$) is time reversible if $P_{X_{t_1},\ldots,X_{t_k}} = P_{X_{t_k},\ldots,X_{t_1}}$, $\forall\, t_1, \ldots, t_k \in T$ and $\forall\, k \in \mathbb{N}$. Chen, Chou and Kuan [24] showed that time reversibility of $(X_t)_{t\in T}$ implies symmetry of the differences $Y_{t,k} = X_t - X_{t-k}$, $\forall\, k \in \mathbb{N}$, about the origin. For that reason time reversibility can be checked with the aid of symmetry tests. Of course, in practice one is not able to test symmetry of $(Y_{t,k})_{k\in\mathbb{N}}$ but only of $(Y_{t,k})_{k=1}^K$ for some fixed K. Several conclusions concerning appropriate models to describe the observed process can be drawn from rejecting the hypothesis of time reversibility. Of course, in this situation the underlying variables are not independent and identically distributed. This outcome of the test also contradicts the underlying variables to form a stationary Gaussian process, cf. Weiss [108].

A great variety of symmetry tests for i.i.d. random variables are available in the literature. Detailed lists of references are provided by Lee [85] as well as by Henze, Klar and Meintanis [71]. Also in the context of time series numerous tests of symmetry have been employed. There are several moment-based tests. Exemplarily, we mention the approach of Ramsey and Rothman [98] that tests whether the third moments vanish and the one of Bai and Ng [7] that is based on the sample skewness coefficient. Obviously, these kinds of tests are inconsistent against asymmetric alternatives whose third moments are zero.

For examples of such distributions see Li and Morris [87] or Ord [95]. Lee [85] built an M test for the composite hypothesis of symmetry around an unknown location parameter. However, a consistency result is not stated. The test for symmetry around the origin of Chen, Chou and Kuan [24] employs the fact that a distribution is symmetric if and only if the imaginary part of its characteristic function $\Im c(t) = \mathbb{E}\sin(t'X)$ vanishes. No moment restrictions on the marginal distributions of the underlying data are required. Again, they did not derive consistency of their test. Fan and Ullah [66] considered a consistent L_2-test for symmetry around the origin based on kernel density estimates with vanishing bandwidths. They assumed the underlying sample to arise from an absolute regular process whose marginal distribution has a bounded and continuous density w.r.t. the Lebesgue measure. In what follows we will derive a consistent characteristic function-based test for symmetry around an unknown center that is asymptotically unbiased against Pitman local alternatives.

5.3.2 The test statistic and its asymptotic behaviour

Suppose that $X_1, \ldots, X_n$ are $\mathbb{R}^d$-valued random variables with distribution P_X and characteristic function c, given by $c(t) = \mathbb{E}e^{it'X_1},\ t \in \mathbb{R}^d$. We establish a consistent test for the problem

$$\mathcal{H}_0: \quad P_{X-\mu_0} = P_{\mu_0 - X} \quad \text{for some } \mu_0 \in \mathbb{R}^d \quad \text{against} \quad \mathcal{H}_1: \quad P_{X-\mu} \neq P_{\mu - X} \quad \forall\, \mu \in \mathbb{R}^d,$$

which can be equivalently expressed in terms of the imaginary part of the characteristic function

$$\mathcal{H}_0: \quad \Im[e^{-it'\mu_0}c(t)] \equiv 0 \quad \text{for some } \mu_0 \in \mathbb{R}^d \quad \text{against} \quad \mathcal{H}_1: \quad \Im[e^{-it'\mu}c(t)] \neq 0 \quad \forall\, \mu \in \mathbb{R}^d.$$

Motivated by the approaches of Feuerverger and Mureika [67], who tested for symmetry around the origin in the case of i.i.d. univariate observations, and Henze, Klar and Meintanis [71], who investigated the composite hypothesis for i.i.d. multivariate random variables, we suggest the following test statistic:

$$\widehat{S}_n := n \int_{\mathbb{R}^d} \left(\Im\left[e^{-it'\widehat{\mu}_n} c_n(t) \right] \right)^2 w(t)\, dt = \int_{\mathbb{R}^d} \left[\frac{1}{\sqrt{n}} \sum_{k=1}^n \sin(t'(X_k - \widehat{\mu}_n)) \right]^2 w(t)\, dt.$$

Here, $\widehat{\mu}_n$ denotes a $\sqrt{n}$-consistent estimator of μ_0. The empirical characteristic function c_n based on the observations $X_1, \ldots, X_n$ is defined through $c_n(t) = n^{-1}\sum_{k=1}^n e^{it'X_k},\ t \in \mathbb{R}^d$. For notational simplicity we refer to c_{μ_0} as the characteristic function of the random variable $X - \mu_0$ in this section. Under weak constraints, the statistic $\widehat{S}_n$ can be approximated by a degenerate V-statistic. We make the following assumptions:

(S1) **(i)** $(X_n)_{n\in\mathbb{N}}$ is a stationary sequence of τ-dependent $\mathbb{R}^d$-valued random variables on some probability space $(\Omega, \mathcal{A}, P)$. Moreover, $\sum_{r=1}^\infty r(\tau_r)^{\delta^2} < \infty$ and $\mathbb{E}\|X_1\|_1^{1+\varepsilon} < \infty$ for all $\varepsilon < 1/(3-2\delta)$ and some $\delta \in (0,1)$.

(ii) The sequence of estimators $\widehat{\mu}_n$ admits the expansion

$$\widehat{\mu}_n - \mu_0 = \frac{1}{n} \sum_{k=1}^{n} l(X_k, \mu_0) + o_P(n^{-1/2}),$$

where $l : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a Lipschitz continuous function satisfying $\mathbb{E}l(X_1, \mu_0) = 0_d$ and $\mathbb{E}\|l(X_1, \mu_0)\|_1^{2\nu} < \infty$ for some $\nu > (2-\delta)/(1-\delta)$.

(iii) The function w is measurable, positive a.s. w.r.t. the Lebesgue measure on $\mathbb{R}^d$ and satisfies $\int_{\mathbb{R}^d} (1 + \|t\|_1^{2\nu})\, w(t)\, dt < \infty$.

Remark 5.2. (i) The assumption of l being Lipschitz continuous can be weakened at the expense of additional moment constraints. However, if one simply uses the arithmetic mean to estimate μ_0, the condition (S1)(ii) is satisfied as long as $\mathbb{E}\|X_1\|_1^{2\nu} < \infty$.

(ii) Note that it does not suffice to postulate strict positivity of w on an interval around the centre of symmetry since there are non-symmetric distributions whose characteristic functions are real-valued in a certain neighbourhood of the origin; see e.g. Ushakov [107], Example 21, Appendix A.

Lemma 5.7. *Suppose that the assumption (S1) holds. Then, under $\mathcal{H}_0$,*

$$\widehat{S}_n - S_n = o_P(1),$$

where $S_n := n \int_{\mathbb{R}^d} \left[n^{-1} \sum_{k=1}^{n} \sin(t'(X_k - \mu_0)) - c_{\mu_0}(t)\, t'\, l(X_k, \mu_0) \right]^2 w(t)\, dt$.

Note that the effect of estimating the unknown parameter is not asymptotically negligible. In contrast to the corresponding test statistic $S_n^{(\mu_0)} := n \int_{\mathbb{R}^d} \left[n^{-1} \sum_{k=1}^{n} \sin(t'(X_k - \mu_0)) \right]^2 w(t)\, dt$ in the case of a known parameter μ_0, we have an extra term $-c_{\mu_0}(t)\, t'\, l(\cdot, \mu_0)$ within the sum here.

In order to verify the assertion above, an auxiliary result concerning variances of sums of τ-dependent observations will be required. Since the behaviour of the test statistic under local alternatives is investigated later, the next result is stated for the more general framework of triangular schemes of random variables.

Lemma 5.8. *Let $(X_{n,k})_{k=1}^n, n \in \mathbb{N}$, be a stationary triangular scheme of $\mathbb{R}^d$-valued random variables such that (A5)(i) holds for some $\delta \in (0,1)$. Suppose that the sequence of functions $g_n : \mathbb{R}^d \to \mathbb{R}$ satisfy $\sup_{n\in\mathbb{N}} \mathbb{E}|g_n(X_{n,1})|^{(2-\delta)/(1-\delta)} < \infty$ and*

$$|g_n(x) - g_n(y)| \le f_n(x, y)\, \|x - y\|_1, \quad \forall\, x, y \in \mathbb{R}^d,$$

where the functions $f_n : \mathbb{R}^{2d} \to \mathbb{R}$, $n \in \mathbb{N}$, fulfil $\sup_{n\in\mathbb{N}} \mathbb{E}\{[f_n(X,Y)]^{1/(1-\delta)}(\|X\|_1 + \|Y\|_1)\} < \infty$ for i.i.d. random variables X, Y with $X \sim P_{X_{n,1}}$. Then there exists a finite constant K such that

$$\sup_{n\in\mathbb{N}} \operatorname{var}\left[\frac{1}{\sqrt{n}} \sum_{k=1}^{n} g_n(X_{n,k}) \right] \le K.$$

Proof. We have

$$\operatorname{var}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{n} g_n(X_{n,k})\right] = \frac{1}{n}\sum_{k=1}^{n}\operatorname{var}(g_n(X_{n,k})) + \frac{2}{n}\sum_{1\le j<k\le n}\operatorname{cov}(g_n(X_{n,j}), g_n(X_{n,k})).$$

Denote by $\widetilde{X}_{n,k}$ a copy of $X_{n,k}$ that is independent of $X_{n,j}$ and that satisfies $\mathbb{E}\|X_{n,k} - \widetilde{X}_{n,k}\|_1 \le \bar{\tau}_{k-j}$. According to equation (2.5) such a random variable exists, at least after enlarging the underlying probability space. Based on this relation we can bound the covariance terms by an iterative application of Hölder's inequality:

$$\begin{aligned}|\operatorname{cov}(g_n(X_{n,j}), g_n(X_{n,k}))| &= |\mathbb{E} g_n(X_{n,j})\,[g_n(X_{n,k}) - g_n(\widetilde{X}_{n,k})]| \\ &\le C(\mathbb{E}|g_n(X_{n,k}) - g_n(\widetilde{X}_{n,k})|)^{\delta} \\ &\le C(\bar{\tau}_{k-j})^{\delta^2}.\end{aligned}$$

Eventually, this leads to

$$\operatorname{var}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{n} g_n(X_{n,k})\right] \le C + \frac{C}{n}\sum_{1\le j<k\le n}(\bar{\tau}_{k-j})^{\delta^2} \le C\left[1 + \sum_{r=1}^{\infty}(\bar{\tau}_r)^{\delta^2}\right] < \infty,$$

which in turn implies the assertion. □

Remark 5.3. If the functions $g_n,\ n\in\mathbb{N}$, in the lemma above are either bounded or Lipschitz continuous, then we only require $\sum_{r=1}^{\infty}(\bar{\tau}_r)^{\delta} < \infty$, which is less restrictive than the summability condition in (A5)(i).

Proof of Lemma 5.7. The approximation will be carried out in two steps. First, define

$$\widetilde{S}_n := \int_{\mathbb{R}^d}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left[\sin(t'(X_k-\mu_0)) - (\widehat{\mu}_n-\mu_0)'t\cos(t'(X_k-\mu_0))\right]\right)^2 w(t)\,dt.$$

<u>Step 1:</u> $\widehat{S}_n - \widetilde{S}_n = o_P(1)$.
The application of an addition formula for trigonometric functions yields

$$\sin(t'(X_k-\widehat{\mu}_n)) = \sin(t'(X_k-\mu_0))\cos(t'(\mu_0-\widehat{\mu}_n)) + \cos(t'(X_k-\mu_0))\sin(t'(\mu_0-\widehat{\mu}_n)).$$

Consequently,

$$\begin{aligned}\widehat{S}_n - \widetilde{S}_n = &\int_{\mathbb{R}^d}\frac{1}{n}\sum_{j,k=1}^{n}\sin(t'(X_j-\mu_0))\sin(t'(X_k-\mu_0))[\cos^2(t'(\mu_0-\widehat{\mu}_n))-1]\,w(t)\,dt \\ &+\int_{\mathbb{R}^d}\frac{2}{n}\sum_{j,k=1}^{n}\Big\{\sin(t'(X_j-\mu_0))\cos(t'(X_k-\mu_0)) \\ &\qquad\times[\sin(t'(\mu_0-\widehat{\mu}_n))\cos(t'(\mu_0-\widehat{\mu}_n)) + t'(\widehat{\mu}_n-\mu_0)]\Big\}\,w(t)\,dt \\ &+\int_{\mathbb{R}^d}\frac{1}{n}\sum_{j,k=1}^{n}\Big\{\cos(t'(X_j-\mu_0))\cos(t'(X_k-\mu_0)) \\ &\qquad\times[\sin^2(t'(\mu_0-\widehat{\mu}_n)) - (t'(\widehat{\mu}_n-\mu_0))^2]\Big\}\,w(t)\,dt \\ =: &\,T_1+T_2+T_3.\end{aligned}$$

Since $\sqrt{n}\,(\widehat{\mu}_n - \mu_0) = n^{-1/2}\sum_{k=1}^n l(X_k,\mu_0) + o_P(1)$, Lemma 5.8 implies $\sqrt{n}\,(\widehat{\mu}_n - \mu_0) = O_p(1)$. This leads to

$$|\cos^2(t'(\mu_0-\widehat{\mu}_n)) - 1| = \left|(\cos(t'(\mu_0-\widehat{\mu}_n)) - 1)(\cos(t'(\mu_0-\widehat{\mu}_n)) + 1)\right| \le \|t\|_1\, o_P(1),$$

which in turn implies $T_1 = o_P(1)\int_{\mathbb{R}^d}[n^{-1/2}\sum_{k=1}^n \sin(t'(X_k-\mu_0))]^2\,\|t\|_1\,w(t)\,dt$. In order to prove that the integral is of order $O_P(1)$, it remains to show that for some $C<\infty$ the inequality

$$\int_{\mathbb{R}^d}\mathbb{E}\left[\frac{1}{\sqrt{n(1+\|t\|_1)}}\sum_{k=1}^n \sin(t'(X_k-\mu_0))\right]^2 \|t\|_1\,(1+\|t\|_1)\,w(t)\,dt < C,\ \forall\, n\in\mathbb{N},$$

holds. Applying Lemma 5.8 with $g_n(x) = \sin(t'(x-\mu_0))/(1+\|t\|_1)$, we obtain $\sup_t \mathbb{E}[(n(1+\|t\|_1))^{-1/2}\sum_{k=1}^n \sin(t'(X_k-\mu_0))]^2 < \infty$. The integrability constraint on the function w finally leads to $T_1 = o_P(1)$.

Furthermore, the equality $2\sin(t'(\mu_0-\widehat{\mu}_n))\cos(t'(\mu_0-\widehat{\mu}_n)) = \sin(2t'(\mu_0-\widehat{\mu}_n))$ and a Taylor expansion of $\sin(2t'(\mu_0-\widehat{\mu}_n))$ in the origin yield

$$\begin{aligned}
|T_2| &\le \int_{\mathbb{R}^d}\left|\sum_{k=1}^n \sin(t'(X_k-\mu_0))\right|\left|2\sin(t'(\mu_0-\widehat{\mu}_n))\cos(t'(\mu_0-\widehat{\mu}_n)) - 2t'(\mu_0-\widehat{\mu}_n)\right| w(t)dt\\
&= \int_{\mathbb{R}^d}\left|\sum_{k=1}^n \sin(t'(X_k-\mu_0))\right|\left|\sin(2t'(\mu_0-\widehat{\mu}_n)) - 2t'(\mu_0-\widehat{\mu}_n)\right| w(t)\,dt\\
&\le 2\int_{\mathbb{R}^d}\left|\sum_{k=1}^n \sin(t'(X_k-\mu_0))\right| [t'(\mu_0-\widehat{\mu}_n)]^2\,w(t)\,dt.
\end{aligned}$$

Employing Lemma 5.8 in the same manner as before, we get $|T_2| = O_P(n^{-1/2})\int_{\mathbb{R}^d}\|t\|_1^2\,(1+\|t\|_1)\,w(t)\,dt = o_P(1)$.

Finally, the quantity T_3 can be approximated by applying the identity $2\sin^2(t'(\mu_0-\hat{\mu}_n)) = 1-\cos(2t'(\mu_0-\widehat{\mu}_n))$ and by Taylor expansion of $\cos(2t'(\mu_0-\widehat{\mu}_n))$ in the origin:

$$\begin{aligned}
|T_3| &\le n\int_{\mathbb{R}^d}\left|\frac{1}{2}\left[1-\cos(2t'(\mu_0-\widehat{\mu}_n))\right] - [t'(\mu_0-\widehat{\mu}_n)]^2\right| w(t)\,dt\\
&\le \frac{n}{6}\int_{\mathbb{R}^d}\left|t'(\mu_0-\widehat{\mu}_n)\right|^3 w(t)\,dt\\
&\le O_P(n^{-1/2})\int_{\mathbb{R}^d}(1+\|t\|_1^3)\,w(t)\,dt,
\end{aligned}$$

which completes the proof of step 1.

<u>*Step 2:*</u> $\widetilde{S}_n - S_n = o_P(1)$.

We split up

$$\begin{aligned}
\widetilde{S}_n - S_n = &\int_{\mathbb{R}^d} \frac{2}{n} \sum_{j,k=1}^{n} \sin(t'(X_j - \mu_0)) \\
&\quad \times [c_{\mu_0}(t)\, l(X_k,\mu_0)'\, t - (\widehat{\mu}_n - \mu_0)'\, t\, \cos(t(X_k - \mu_0))]\, w(t)\, dt \\
&+ \int_{\mathbb{R}^d} \frac{1}{n} \sum_{j,k=1}^{n} \Big[\cos(t'(X_j - \mu_0))\, \cos(t'(X_k - \mu_0))\, [t'(\widehat{\mu}_n - \mu_0)]^2 \\
&\quad - c_{\mu_0}^2(t)\, l(X_j,\mu_0)'\, t\, l(X_k,\mu_0)' t \Big]\, w(t)\, dt \\
= &\, R_1 + R_2.
\end{aligned}$$

Concerning the first summand one gets

$$\begin{aligned}
|R_1| \le &\int_{\mathbb{R}^d} \Big| \frac{2}{\sqrt{n}} \sum_{j=1}^{n} \sin(t'(X_j - \mu_0)) \Big| |c_{\mu_0}(t)| \Big| \frac{1}{\sqrt{n}} \sum_{k=1}^{n} l(X_k,\mu_0)' t - \sqrt{n}(\widehat{\mu}_n - \mu_0)' t \Big|\, w(t)\, dt \\
&+ \int_{\mathbb{R}^d} \Big| \frac{2}{\sqrt{n}} \sum_{j=1}^{n} \sin(t'(X_j - \mu_0)) \Big| \Big| c_{\mu_0}(t) - \frac{1}{n} \sum_{k=1}^{n} \cos(t'(X_k - \mu_0)) \Big| \\
&\times |\sqrt{n}(\widehat{\mu}_n - \mu_0)' t|\, w(t)\, dt \\
\le &\, o_P(1) \int_{\mathbb{R}^d} \Big| \frac{2}{\sqrt{n}} \sum_{j=1}^{n} \sin(t'(X_j - \mu_0)) \Big| \|t\|_1\, w(t)\, dt + O_P(1) \\
&\times \int_{\mathbb{R}^d} \Big| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sin(t'(X_j - \mu_0)) \Big| \Big| c_{\mu_0}(t) - \frac{1}{n} \sum_{k=1}^{n} \cos(t'(X_k - \mu_0)) \Big| \|t\|_1\, w(t)\, dt.
\end{aligned}$$

According to Lemma 5.8, the first summand vanishes asymptotically. Hence, it remains to show that

$$\mathbb{E} \int_{\mathbb{R}^d} \Big| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sin(t'(X_j - \mu_0)) \Big| \Big| c_{\mu_0}(t) - \frac{1}{n} \sum_{k=1}^{n} \cos(t'(X_k - \mu_0)) \Big| \|t\|_1\, w(t)\, dt$$

tends to zero as $n \to \infty$. Similar to our previous investigations, this expression can be bounded under $\mathcal{H}_0$ by

$$\begin{aligned}
&C \int_{\mathbb{R}^d} \left(\mathbb{E} \left[c_{\mu_0}(t) - \frac{1}{n} \sum_{k=1}^{n} \cos(t'(X_k - \mu_0)) \right]^2 \right)^{1/2} \sqrt{1 + \|t\|_1}\, \|t\|_1\, w(t)\, dt \\
&\le O\left(n^{-1/2}\right) \int_{\mathbb{R}^d} (1 + \|t\|_1) \|t\|_1\, w(t)\, dt.
\end{aligned}$$

Finally, concerning R_2 we obtain

$$\begin{aligned}
|R_2| &= o_P(1) + \Big| \int_{\mathbb{R}^d} \frac{1}{n} \sum_{i,j=1}^{n} l(X_i,\mu_0)'t\, l(X_j,\mu_0)'t \\
&\quad \times \Big[\frac{1}{n^2} \sum_{k,m=1}^{n} \cos(t'(X_k-\mu_0))\cos(t'(X_m-\mu_0)) - c_{\mu_0}^2(t)\Big]\, w(t)\, dt\Big| \\
&= o_P(1) + O_P(1) \int_{\mathbb{R}^d} \Big| \Big(\frac{1}{n}\sum_{k=1}^{n} \cos(t'(X_k-\mu_0))\Big)^2 - c_{\mu_0}^2(t)\Big|\, \|t\|_1^2\, w(t)\, dt \\
&= o_P(1) + O_P(1) \int_{\mathbb{R}^d} \Big|\frac{1}{n}\sum_{k=1}^{n} \cos(t'(X_k-\mu_0)) - c_{\mu_0}(t)\Big|\, \|t\|_1^2\, w(t)\, dt \\
&= o_P(1)
\end{aligned}$$

with the aid of Lemma 5.8. □

Obviously,

$$\begin{aligned}
S_n = \frac{1}{n} \sum_{j,k=1}^{n} \int_{\mathbb{R}^d} &\big[\sin(t'(X_j-\mu_0)) - c_{\mu_0}(t)\, t'\, l(X_j,\mu_0)\big] \\
&\times \big[\sin(t'(X_k-\mu_0)) - c_{\mu_0}(t)\, t'\, l(X_k,\mu_0)\big]\; w(t)\, dt
\end{aligned}$$

is a degenerate V-statistic (multiplied with n) with a symmetric and continuous kernel. The assumption (S1) assures the validity of the conditions (A1), (A2), and (A4) of Chapter 3. Therefore, Theorem 3.2 can be applied in order to derive the asymptotic distribution of our test statistic.

Theorem 5.1. *Suppose that the assumption (S1) holds. Then, under the null hypothesis,*

$$\widehat{S}_n \xrightarrow{d} Z_S + \mathbb{E}h_{\mu_0}(X_1,X_1).$$

Here, the distribution of $Z_S + \mathbb{E}h_{\mu_0}(X_1,X_1)$ is the limit distribution of $n\,V_n$ defined in Theorem 3.2, where V_n denotes a V-statistic based on the underlying sample and whose kernel function is given by

$$h_{\mu_0}(x,y) := \int_{\mathbb{R}^d} [\sin(t'(x-\mu_0)) - c_{\mu_0}(t)t'l(x,\mu_0)][\sin(t'(y-\mu_0)) - c_{\mu_0}(t)t'l(y,\mu_0)]w(t)dt.$$

Proof. Due to Lemma 5.7 it suffices to show that $S_n \xrightarrow{d} Z_S + \mathbb{E}h_{\mu_0}(X_1,X_1)$. To prove this, we verify the validity of the prerequisites of Theorem 3.2 with $S_n = n\,V_n$ where V_n is the V-statistic with kernel h_{μ_0}. The assumption (S1)(i) implies the condition (A1). Obviously, h_{μ_0} is symmetric, degenerate under the null hypothesis and satisfies $\sup_{k\in\mathbb{N}} \mathbb{E}|h_{\mu_0}(X_1,X_{1+k})|^{\nu} + \mathbb{E}|h_{\mu_0}(X_1,\widetilde{X}_1)|^{\nu} < \infty$, where $\widetilde{X}_1$ is an i.i.d. copy of the random variable X_1. Thus, also the assumption (A2) holds and it remains to check (A4).

The function h_{μ_0} has the following continuity property:

$$\begin{aligned}|h_{\mu_0}(x,y) - h_{\mu_0}(\bar{x},\bar{y})| &\le C\,(1 + \|l(x,\mu_0)\|_1 + \|l(y,\mu_0)\|_1 + \|l(\bar{x},\mu_0)\|_1 + \|l(\bar{y},\mu_0)\|_1) \\ &\qquad \times [\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1] \\ &=: f_{\mu_0}(x,\bar{x},y,\bar{y})\,[\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1].\end{aligned} \tag{5.16}$$

Hence, due to stationarity, the Lipschitz continuity of l, and the moment constraints on l and X_1, the application of Hölder's inequality implies

$$\begin{aligned}&\sup_{k_1,\ldots,k_5\in\mathbb{N}} \mathbb{E}\left(\max_{a_1,a_2\in[-A,A]^d}[f_{\mu_0}(Y_{k_1},Y_{k_2}+a_1,Y_{k_3},Y_{k_4}+a_2)]^\eta \|Y_{k_5}\|_1\right) \\ &\le C\left[1 + \sup_{k_1,k_2\in\mathbb{N}} \mathbb{E}\left(\max_{a_1\in[-A,A]^d} \|l(Y_{k_1}+a_1,\mu_0)\|_1^\eta\, \|Y_{k_2}\|_1\right)\right] \\ &\le C\left[1 + \sup_{k_1,k_2\in\mathbb{N}} \mathbb{E}[\|l(Y_{k_1},\mu_0)\|_1 + A\,d]^\eta\, \|Y_{k_2}\|_1\right] \\ &\le C\left[1 + \left(\mathbb{E}\|l(X_1,\mu_0)\|_1^{2\nu}\right)^{\eta/(2\nu)} \left(\mathbb{E}\|X_1\|_1^{2\nu/(2\nu-\eta)}\right)^{(2\nu-\eta)/(2\nu)}\right] \\ &< \infty\end{aligned}$$

for $\eta := 1/(1-\delta)$, some $A > 0$, and for any $(Y'_{k_1},\ldots,Y'_{k_5})'$ consisting of independent subvectors $(Y'_{k_{j_1(m)}},\ldots,Y'_{k_{j_l(m)}})' \stackrel{d}{=} (X'_{k_{j_1(m)}},\ldots,X'_{k_{j_l(m)}})'$, $l,m = 1,\ldots,5$. To this end, also note that $2\nu/(2\nu-\eta) < 1 + (3-2\delta)^{-1}$ which actually implies $\mathbb{E}\|X_1\|_1^{2\nu/(2\nu-\eta)} < \infty$ by virtue of (S1)(i). Consequently, all prerequisites of Theorem 3.2 are satisfied. □

Nevertheless, the limit distribution of the test statistic is of a complicated structure. Therefore, problems arise as soon as (asymptotic) critical values of the test have to be determined. In the following subsection we propose the application of certain bootstrap methods in order to circumvent these difficulties.

Next, the behaviour of the test statistic under fixed alternatives is studied.

Lemma 5.9. *Suppose that (S1)(i) and (iii) hold. Moreover, assume that there exists a constant $\mu_0 \in \mathbb{R}^d$ such that $\widehat{\mu}_n \xrightarrow{P} \mu_0$. Then, under $\mathcal{H}_1$,*

$$P\left(\widehat{S}_n > K\right) \underset{n\to\infty}{\longrightarrow} 1, \quad \forall\, K \in \mathbb{R}.$$

Proof. Under $\mathcal{H}_1$ and the assumptions concerning the weight function w, $\int_{\mathbb{R}^d}[\Im(c_{\mu_0}(t))]^2\, w(t)\, dt > 0$ holds true. Thus, it is sufficient to show that $P(|n^{-1}\widehat{S}_n - \int_{\mathbb{R}^d}[\Im(c_{\mu_0}(t))]^2\, w(t)\, dt| \ge \varepsilon) \longrightarrow_{n\to\infty} 0$, $\forall\,\varepsilon > 0$, in order to verify the claim. For all $\varepsilon > 0$ there exists a $T \in (0,\infty)$ such that $\int_{\mathbb{R}^d\setminus[-T,T]^d} w(t)\, dt < \varepsilon/2$ according to the integrability assumption on w. Hence, it remains to verify

$$P\left(\int_{[-T,T]^d} \left|\Im(e^{it'\widehat{\mu}_n} c_n(t)) - \Im(c_{\mu_0}(t))\right| w(t)dt > \frac{\varepsilon}{4}\right) \underset{n\to\infty}{\longrightarrow} 0. \tag{5.17}$$

Clearly, $\int_{[-T,T]^d}(\cdot)(t)\,w(t)\,dt$ is a continuous mapping from $C([-T,T]^d)$ to $\mathbb{R}^+$, both endowed with the corresponding uniform metrics. The continuous mapping theorem implies (5.17) if $(n^{-1}\sum_{j=1}^n \sin[t'(X_j-\widehat{\mu}_n)])_{t\in[-T,T]^d} \xrightarrow{P} (\Im[c_{\mu_0}(t)])_{t\in[-T,T]^d}$ holds true. To this end, we first show pointwise convergence $n^{-1}\sum_{j=1}^n \sin[t'(X_j-\widehat{\mu}_n)] \xrightarrow{P} \Im[c_{\mu_0}(t)]$ for any fixed $t\in\mathbb{R}^d$. Afterwards it remains to prove that

$$\begin{aligned}\lim_{\delta\to 0}\limsup_{n\to\infty} P\Big(\sup_{\|s-t\|_1<\delta}\Big|\frac{1}{n}\sum_{j=1}^n\Big\{&\sin[s'(X_j-\widehat{\mu}_n)]-\Im(c_{\mu_0}(s))\\ &-\sin[t'(X_j-\widehat{\mu}_n)]+\Im(c_{\mu_0}(t))\Big\}\Big|\geq\varepsilon\Big)=0\end{aligned}\tag{5.18}$$

for all $\varepsilon>0$. In order to derive pointwise convergence, note that

$$\begin{aligned}&\left|\frac{1}{n}\sum_{j=1}^n\sin[t'(X_j-\widehat{\mu}_n)]-\Im[c_{\mu_0}(t)]\right|\\ &=\left|\frac{1}{n}\sum_{j=1}^n\big(\sin[t'(X_j-\mu_0)]\cos[t'(\mu_0-\widehat{\mu}_n)]+\sin[t'(\mu_0-\widehat{\mu}_n)]\cos[t'(X_j-\mu_0)]\big)-\Im[c_{\mu_0}(t)]\right|\\ &\leq\left|\frac{1}{n}\sum_{j=1}^n\sin[t'(X_j-\mu_0)]-\Im[c_{\mu_0}(t)]\right|+\|t\|_1\,o_P(1).\end{aligned}$$

Due to Lemma 2.3 the remaining sum converges to zero in probability. For proving the relation (5.18), it suffices to verify

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\left(\sup_{\|s-t\|_1<\delta}\left|\frac{1}{n}\sum_{j=1}^n\Big\{\sin[s'(X_j-\widehat{\mu}_n)]-\sin[t'(X_j-\widehat{\mu}_n)]\Big\}\right|\geq\frac{\varepsilon}{2}\right)=0$$

according to the uniform continuity of the characteristic function. We obtain

$$\begin{aligned}&\lim_{\delta\to 0}\limsup_{n\to\infty} P\left(\sup_{\|s-t\|_1<\delta}\left|\frac{1}{n}\sum_{j=1}^n\Big\{\sin[s'(X_j-\widehat{\mu}_n)]-\sin[t'(X_j-\widehat{\mu}_n)]\Big\}\right|\geq\frac{\varepsilon}{2}\right)\\ &\leq\lim_{\delta\to 0}\limsup_{n\to\infty} P\left(\frac{\delta}{n}\sum_{j=1}^n\|X_j-\widehat{\mu}_n\|_1\geq\frac{\varepsilon}{2}\right)\\ &\leq\lim_{\delta\to 0}\limsup_{n\to\infty} P\left(\delta\,\|\widehat{\mu}_n-\mu_0\|_1\geq\frac{\varepsilon}{4}\right)+\lim_{\delta\to 0}\limsup_{n\to\infty} P\left(\frac{\delta}{n}\sum_{j=1}^n\|X_j-\mu_0\|_1\geq\frac{\varepsilon}{4}\right)\\ &=0.\end{aligned}$$

The latter probability vanishes asymptotically since $n^{-1}\sum_{j=1}^n\|X_j-\mu_0\|_1 \xrightarrow{P} \mathbb{E}\|X_1-\mu_0\|_1$ under (S1)(i) by virtue of Lemma 2.3. $\square$

This subsection is concluded by an assertion concerning the asymptotics of the test statistic under Pitman local alternatives. More precisely, we consider alternatives

$$\mathcal{H}_{1,n}:\frac{dP_{X_{n,1}}}{dP_{X_1}}=1+\frac{g_n}{\sqrt{n}}\text{ with }\|g_n-g\|_\infty\underset{\mathrm{n}\to\infty}{\longrightarrow}0\text{ and }\int_{\mathbb{R}^d}\sin(t'(x-\mu_0))g(x)P_X(dx)\neq 0,$$

where g is assumed to be a measurable bounded function. The corresponding test statistic is denoted by $\widehat{S}_{n,n,}$, i.e.

$$\widehat{S}_{n,n} := \int_{\mathbb{R}^d} \left[\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sin(t'(X_{n,k} - \widehat{\mu}_{n,n})) \right]^2 w(t)\, dt,$$

Here, $\widehat{\mu}_{n,n}$ denotes an estimator of the location parameter μ_0 based on the observations $(X_{n,k})_{k=1}^n$. In order to derive the limit distribution of the statistic above, one shows that it can be approximated by a statistic Proposition 5.1 can be applied to. Moreover, we intend to study the asymptotic behaviour of a bootstrap-based test under this kind of local alternatives in the next subsection. To verify that the procedure is at least asymptotically unbiased, the validity of

$$\liminf_{n\to\infty} \left[P(\widehat{S}_{n,n} > x) - P(\widehat{S}_n > x) \right] \geq 0, \quad \forall\, x \in \mathbb{R},$$

can be invoked. This assertion follows from Lemma 5.2 if the matrices $\Delta_c^{(K,L)}$ are positive semidefinite for all $K, L \in \mathbb{N}$ and $c \geq c_0$ for some $c_0 \in \mathbb{R}^+$. Indeed, these matrices are easily proved to be positive semidefinite if the kernel truncation method that is applied to determine these matrices preserves the structure of the kernel function that separates the involved variables, i.e. if $h_c(x,y) = \int_{\mathbb{R}^d} h_c^{(1)}(x,t)\, h_c^{(1)}(y,t)\, w(t)\, dt$ for certain functions $h_c^{(1)}$. Unfortunately, the kernel truncation method we used so far does not carry over the variable separating form of the kernel. In order to prove the subsequent assertion, we have to provide an alternative way of truncating the kernel function to obtain the desirable structure.

Proposition 5.3. *Suppose that the conditions (A5) and (S1) are satisfied. Additionally, assume that the sequence of estimators $\widehat{\mu}_{n,n}$ admits the expansion $\widehat{\mu}_{n,n} - \mu_0 = n^{-1} \sum_{k=1}^n l(X_{n,k}, \mu_0) + o_P(n^{-1/2})$. Then, as $n \to \infty$,*

$$\widehat{S}_{n,n} \xrightarrow{d} Z_{S,loc} + \mathbb{E} h_{\mu_0}(X_1, X_1).$$

Here, the distribution of $Z_{S,loc} + \mathbb{E} h_{\mu_0}(X_1, X_1)$ is the limit distribution of $n\, V_{n,n}$ defined in Proposition 5.1, where $V_{n,n}$ denotes the V-statistic based on the underlying sample and whose kernel function is given by h_{μ_0}. If additionally $\operatorname{var}(Z_S) > 0$*, then*

$$\liminf_{n\to\infty} \left[P(\widehat{S}_{n,n} > x) - P(\widehat{S}_n > x) \right] \geq 0, \quad \forall\, x \in \mathbb{R}.$$

Proof. We assume w.l.o.g. that $\sup_{n\in\mathbb{N}} \|g_n\|_\infty < \infty$.
Step 1: $n\, \widehat{S}_{n,n} \xrightarrow{d} Z_{S,loc} + \mathbb{E} h_{\mu_0}(X_1, X_1)$.
First, we show that $\widehat{S}_{n,n}$ can be approximated by the V-statistic

$$S_{n,n} := n \int_{\mathbb{R}^d} \left[\frac{1}{n} \sum_{k=1}^{n} \sin(t'(X_{n,k} - \mu_0)) - c_{\mu_0}(t)\, t'\, l(X_{n,k}, \mu_0) \right]^2 w(t)\, dt$$

whose kernel degenerates under P_{X_1}. Afterwards, Proposition 5.1 is applied to derive the limit of the statistics $S_{n,n}$.

Here, one can proceed as in the proof of Lemma 5.7. The introduction of

$$\widetilde{S}_{n,n} := \int_{\mathbb{R}^d} \left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sin(t'(X_{n,k} - \mu_0)) - (\widehat{\mu}_{n,n} - \mu_0)' \, t \, \cos(t'(X_{n,k} - \mu_0)) \right)^2 w(t) \, dt$$

yields $\widehat{S}_{n,n} - \widetilde{S}_{n,n} = o_P(1)$ in complete analogy to step 1 of the proof of Lemma 5.7 if $\sup_{t \in \mathbb{R}^d} \mathbb{E}\left[(n(1 + \|t\|_1))^{-1/2} \sum_{k=1}^{n} \sin(t'(X_{n,k} - \mu_0)) \right]^2 = O(1)$. To verify this, we split up

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{\sqrt{n(1 + \|t\|_1)}} \sum_{k=1}^{n} \sin(t'(X_{n,k} - \mu_0)) \right]^2 \\
&\le 2\,\mathbb{E}\left[\frac{1}{\sqrt{n(1 + \|t\|_1)}} \sum_{k=1}^{n} \left\{ \sin(t'(X_{n,k} - \mu_0)) - \mathbb{E}\sin(t'(X_{n,1} - \mu_0)) \right\} \right]^2 \\
&\quad + \frac{2}{1 + \|t\|_1} \left(\int_{\mathbb{R}^d} \sin(t'(x - \mu_0)) \, g_n(x) \, dP_{X_1}(x) \right)^2 .
\end{aligned}$$

The latter summand obviously has the order $O(1)$. Moreover, we get the desired size of the first summand due to (A5) and Lemma 5.8.

In order to show that $\widetilde{S}_{n,n} - S_{n,n} = o_P(1)$, one can continue as in step 2 of the proof of Lemma 5.7. Merely the consideration of $[n^2(1 + \|t\|_1)]^{-1}(\mathbb{E} \sum_{k=1}^{n} [\cos(t'(X_{n,k} - \mu_0)) - c_{\mu_0}(t)])^2$ has to be slightly modified. For deriving that this expression is of order $o(1)$, the estimation

$$\begin{aligned}
&\frac{1}{n^2(1 + \|t\|_1)} \, \mathbb{E}\left(\sum_{k=1}^{n} \left[\cos(t'(X_{n,k} - \mu_0)) - c_{\mu_0}(t) \right] \right)^2 \\
&\le o(1) + 2\,\mathbb{E}\left(\frac{1}{n} \sum_{k=1}^{n} \left[\cos(t'(X_{n,k} - \mu_0)) - \mathbb{E}\cos(t'(X_{n,1} - \mu_0)) \right] \right)^2
\end{aligned}$$

can be used, where the remaining term vanishes asymptotically due to the dependence structure of $(X_{n,k})_{k=1}^{n}$, $n \in \mathbb{N}$, and Lemma 5.8. The required order $O_P(1)$ of $n^{-1/2} \sum_{k=1}^{n} l(X_{n,k}, \theta_0)$ can be derived by applying Lemma 5.8 to $n^{-1/2} \sum_{k=1}^{n} \left[l_i(X_{n,k}, \theta_0) - \int_{\mathbb{R}^d} l_i(x, \theta_0) P_{X_{n,1}}(dx) \right]$ and employing the fact that $\sqrt{n} \int_{\mathbb{R}^d} l_i(x, \theta_0) P_{X_{n,1}}(dx) = O(1)$, $i = 1, \ldots, p$. Thus, $\widehat{S}_{n,n} - S_{n,n} = o_P(1)$.

If we are able to verify validity of the condition (A6), Proposition 5.1 can be invoked to determine the limit distribution of $S_{n,n}$ and thus of $\widehat{S}_{n,n}$. The assumption (A6)(i) trivially holds. Moreover,

$$\begin{aligned}
&\sup_{\substack{1 \le k < n, \\ n \in \mathbb{N}}} \mathbb{E}|h(X_{n,1}, X_{n,1+k})|^{\nu} \\
&\le \sup_{\substack{1 \le k < n, \\ n \in \mathbb{N}}} \mathbb{E}\left(\int_{R^d} (1 + \|t\|_1 \, \|l(X_{n,1}, \theta_0)\|_1) \, (1 + \|t\|_1 \, \|l(X_{n,1+k}, \theta_0)\|_1) \, w(t) \, dt \right)^{\nu}
\end{aligned}$$

is bounded since $\mathbb{E}\|l(X_{n,1},\theta_0)\|_1^{2\nu} \leq (1+\|g_n\|_\infty)\mathbb{E}\|l(X_1,\theta_0)\|_1^{2\nu} < \infty$ which implies that (A6)(ii) is satisfied. According to the definition of the function $f = f_{\mu_0}$ in (5.16), the condition (A6)(iii) follows, too. Hence, all prerequisites of Proposition 5.1 are fulfilled and

$$\widehat{S}_{n,n} \xrightarrow{d} Z_{S,loc} + \mathbb{E}h_{\mu_0}(X_1,X_1).$$

Step 2: $\liminf_{n\to\infty}\left[P(\widehat{S}_{n,n} > x) - P(\widehat{S}_n > x)\right] \geq 0, \quad \forall\, x \in \mathbb{R}.$
We will apply Lemma 5.3 to prove that $\liminf_{n\to\infty}\left[P(S_{n,n} > x) - P(S_n > x)\right] \geq 0$ for all $x \in \mathbb{R}$ which then implies the second assertion of this proposition. Thus, it remains to verify that the matrices of wavelet coefficients $\Delta_c^{(K,L)}$, defined in Lemma 5.2, are positive semidefinite for all $K, L \in \mathbb{N}$ and $c \in \mathbb{R}^+$. As already indicated, we suggest to use a kernel truncation method that differs from (3.4), namely a method that is adjusted to the special structure of the underlying kernel. Instead of directly truncating the function h we refer to the function h_1, given by $h_1(x,t) := \sin(t'(x-\mu_0)) - c_{\mu_0}(t)\, t'\, l(x,\mu_0)$. Based on $g_c(t) := \max_{x\in[-c,c]^d}|h_1(x,t)|$, truncated versions of h_1,

$$h_1^{(c)}(x,t) := \begin{cases} h_1(x,t) & \text{for } |h_1(x,t)| \leq g_c(t), \\ -g_c(t) & \text{for } h_1(x,t) < -g_c(t), \\ g_c(t) & \text{for } h_1(x,t) > g_c(t) \end{cases}$$

are introduced. The new truncated versions $(\bar{h}_c)_c$ of the kernel itself are defined through

$$\begin{aligned}\bar{h}_c(x,y) := \int_{\mathbb{R}^d} &\left[h_1^{(c)}(x,t) - \int_{\mathbb{R}^d} h_1^{(c)}(z,t)P_{X_1}(dz)\right] \\ &\times \left[h_1^{(c)}(y,t) - \int_{\mathbb{R}^d} h_1^{(c)}(z,t)P_{X_1}(dz)\right] w(t)\,dt.\end{aligned}$$

Assume that we are able to verify

$$\limsup_{n\to\infty}\mathbb{E}\left(\frac{1}{n}\sum_{\substack{j,k=1\\ j\neq k}}^{n}\left[h_{\mu_0}(X_j,X_k) - \bar{h}_c(X_j,X_k)\right]\right)^2 \underset{c\to\infty}{\longrightarrow} 0, \tag{5.19}$$

$$\limsup_{n\to\infty}\mathbb{E}\left(\frac{1}{n}\sum_{\substack{j,k=1\\ j\neq k}}^{n}\left[h_{\mu_0}(X_{n,j},X_{n,k}) - \bar{h}_c(X_{n,j},X_{n,k})\right]\right)^2 \underset{c\to\infty}{\longrightarrow} 0, \tag{5.20}$$

and that moreover $|\bar{h}_c(x,y) - \bar{h}_c(\bar{x},\bar{y})| \leq \bar{f}(x,\bar{x},y,\bar{y})[\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1]$, $\forall\, c \in \mathbb{R}_+$, holds for some symmetric continuous function $\bar{f} : \mathbb{R}^{4d} \to \mathbb{R}$ with

$$\sup_{1\leq k_1,\ldots,k_5\leq n}\mathbb{E}\left(\max_{a_1,a_2\in[-A,A]^d}[\bar{f}(Y_{n,k_1},Y_{n,k_2}+a_1,Y_{n,k_3},Y_{n,k_4}+a_2)]^\eta\|Y_{n,k_5}\|_1\right) < \infty, \tag{5.21}$$

where $Y_{n,k_1},\ldots,Y_{n,k_5}$ are defined as in (A6). Then the limit variables Z_S and $Z_{S,loc}$ have alternative representations, which originate from the substitution of $\alpha_{J(L);k_1,k_2}^{(c)}$ by

$\bar{\alpha}^{(c)}_{J(L);k_1,k_2} := \iint_{\mathbb{R}^d\times\mathbb{R}^d} \bar{h}_c(x,y)\, \Phi_{J(L),k_1}(x)\, \Phi_{J(L),k_2}(y)\, dx\, dy$ in the formulas (5.6) and (5.7), respectively. Of course, the assertions of Lemma 5.2 and thus of Lemma 5.3 remain valid if we plug in the counterpart $\bar{\Delta}_c^{(K,L)}$ of $\Delta_c^{(K,L)}$ containing the components $\bar{\alpha}^{(c)}_{J;k_1,k_2}$ instead of the respective coefficients $\alpha^{(c)}_{J;k_1,k_2}$. Indeed, the matrix $\bar{\Delta}_c^{(K,L)}$ is positive semidefinite, since for arbitrary $z = (z_1, \ldots, z_{(2K+1)^d})' \in \mathbb{R}^{(2K+1)^d}$ we have

$$z'\bar{\Delta}_c^{(K,L)} z = \int_{\mathbb{R}^d} \left(\sum_{k=1}^{(2K+1)^d} z_k \int_{\mathbb{R}^d} \Phi_{J(L),k}(x)[h_1^{(c)}(x,t) - \mathbb{E}h_1^{(c)}(X_1,t)]\, dx \right)^2 w(t)\, dt \geq 0.$$

Consequently, the proof is complete if we are able to verify (5.19), (5.20), and (5.21). The smoothness of $\bar{h}_c$ can be described by

$$\left|\bar{h}_c(x,y) - \bar{h}_c(x,y)\right| \leq C f_{\mu_0}(x,\bar{x},y,\bar{y})\, [\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1]\,,$$

where f_{μ_0} is defined in the proof of Theorem 5.1. This immediately yields (5.21). In order to prove the validity of (5.19) and (5.20), it is additionally required that $\sup_c \sup_k [\mathbb{E}|\bar{h}_c(X_1, X_{1+k})|^\nu + \mathbb{E}|\bar{h}_c(X_{n,1}, X_{n,1+k})|^\nu] + \mathbb{E}|\bar{h}_c(X_1, \widetilde{X}_1)|^\nu < \infty$, where $\widetilde{X}_1$ denotes an independent copy of X_1, cf. Remark 3.3. Finally, this moment constraint follows from the moment assumption on the function l. $\square$

5.3.3 A consistent bootstrap method

Theorem 5.1 suggests to reject the hypothesis of symmetry at asymptotic significance level α if $\widehat{S}_n > t_\alpha$. Here, t_α denotes the $(1-\alpha)$-quantile of the distribution of $Z_S + \mathbb{E}h_{\mu_0}(X_1, X_1)$. Obviously, the latter random variable depends on the unknown parameter in a complicated way and thus the asymptotic critical values can hardly be derived analytically or be tabulated. Therefore, we consider a parametric bootstrap procedure to approximate these quantities.

To this end, we make the following assumptions concerning the bootstrap sample $(X_k^*)_k$ and the bootstrap estimator $\mu_n^*(X_1^*, \ldots, X_n^*)$ of the location parameter:

(S2) **(i)** The condition (A1*) holds with $\sum_{r=1}^\infty r(\bar{\tau}_r)^{\delta^2}$. Moreover, $\mathbb{E}^*(\|X_1^*\|_1^{1+\varepsilon}) = O_P(1)$ for all $\varepsilon < 1/(3-2\delta)$.

(ii) The sequence of estimators $\widehat{\mu}_n^*$ admits the expansion

$$\widehat{\mu}_n^* - \widehat{\mu}_n = \frac{1}{n}\sum_{k=1}^n l(X_k^*, \widehat{\mu}_n) + o_{P^*}(n^{-1/2}),$$

where $l : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a Lipschitz continuous function satisfying $\mathbb{E}^* l(X_1^*, \widehat{\mu}_n) = 0_d$ and $\mathbb{E}^*\|l(X_1^*, \widehat{\mu}_n)\|_1^{2\bar{\nu}} = O_P(1)$ for some $\bar{\nu} > (2-\delta)/(1-\delta)$.

Intuitively, one might consider to use the bootstrap counterpart of $\widehat{S}_n$ as the bootstrap test statistic. Unfortunately, it turns out, that the bootstrap version of the approximating

V-statistic S_n is no longer degenerate in general. This is however not surprising since a similar problem occurs when Efron's bootstrap is applied to V-type statistics of i.i.d. data. Dehling and Mikosch [40] proposed to degenerate the kernel on the bootstrap side artificially to overcome this difficulty. We will proceed analogously here, cf. the proof of Proposition 5.6, and propose the following bootstrap algorithm:

1. Determine $\widehat{\theta}_n$.

2. Generate $X_1^*, \ldots, X_n^*$ such that (S2)(i) holds.

3. Determine $\widehat{\mu}_n^*$ such that (S2)(ii) is satisfied.

4. Compute the bootstrap version of our test statistic:

$$\widehat{S}_n^* := \frac{1}{n} \sum_{j,k=1}^{n} \int_{\mathbb{R}^d} \left\{\sin(t'(Y_j^* - \widehat{\mu}_n^*)) - \Im[c^*(t)]\right\} \left\{\sin(t'(Y_k^* - \widehat{\mu}_n^*)) - \Im[c^*(t)]\right\} w(t)\, dt.$$

   Here, c^* is the characteristic function of $X_1^* - \widehat{\mu}_n$, conditionally on $X_1, \ldots, X_n$.

5. Define the critical value t_α^* as the $(1-\alpha)$-quantile of the (conditional) distribution of $\widehat{S}_n^*$. Reject $\mathcal{H}_0$ if $\widehat{S}_n > t_\alpha^*$.
   (In practice, steps 1. to 4. will be repeated B times, for some large B. The critical value will then be approximated by the $(1-\alpha)$-quantile of the empirical distribution associated with $\widehat{S}_{n,1}^*, \ldots, \widehat{S}_{n,B}^*$.)

Remark 5.4. The bootstrap characteristic function c^* is often unknown. However, this function can always be approximated arbitrarily well by simulation.

The subsequent result is an immediate consequence of Proposition 4.1.

Proposition 5.4. *Suppose that the assumptions (S1) and (S2) hold and that $\widehat{S}_n^*$ is generated via the aforementioned algorithm. Then, under $\mathcal{H}_0$,*

$$\widehat{S}_n^* \xrightarrow{d} Z_S + \mathbb{E}h_{\mu_0}(X_1, X_1) \quad \textit{in probability,}$$

as $n \to \infty$, where Z_S and h_{μ_0} are defined as in Theorem 5.1. Moreover, if $\operatorname{var}(Z_S) > 0$, *then*

$$\sup_{-\infty<x<\infty} \left| P(\widehat{S}_n^* \leq x | X_1, \ldots, X_n) - P(\widehat{S}_n \leq x) \right| \xrightarrow{P} 0.$$

Proof. Proceeding as in the proof of Lemma 5.7, we obtain

$$\begin{aligned}
\widehat{S}_n^* &= \int_{\mathbb{R}^d} \Big(\frac{1}{\sqrt{n}} \sum_{k=1}^n \Big[\big\{ \sin(t'(X_k^* - \widehat{\mu}_n)) - \Im[c^*(t)] \big\} \cos(t'(\widehat{\mu}_n - \widehat{\mu}_n^*)) \\
&\quad + \cos(t'(X_k^* - \widehat{\mu}_n)) \sin(t'(\widehat{\mu}_n - \widehat{\mu}_n^*)) + \Im[c^*(t)](\cos(t'(\widehat{\mu}_n - \widehat{\mu}_n^*)) - 1) \Big] \Big)^2 w(t)\, dt \\
&= \int_{\mathbb{R}^d} \Big(\frac{1}{\sqrt{n}} \sum_{k=1}^n \Big[\big\{ \sin(t'(X_k^* - \widehat{\mu}_n)) - \Im[c^*(t)] \big\} \cos(t'(\widehat{\mu}_n - \widehat{\mu}_n^*)) \\
&\quad + \cos(t'(X_k^* - \widehat{\mu}_n)) \sin(t'(\widehat{\mu}_n - \widehat{\mu}_n^*)) \Big] \Big)^2 w(t)\, dt + o_{P^*}(1) \\
&= \int_{\mathbb{R}^d} \Big(\frac{1}{\sqrt{n}} \sum_{k=1}^n \sin(t'(X_k^* - \widehat{\mu}_n)) - \Im[c^*(t)] - (\widehat{\mu}_n^* - \widehat{\mu}_n)' t \cos(t'(X_k^* - \widehat{\mu}_n)) \Big)^2 w(t)\, dt \\
&\quad + o_{P^*}(1) \\
&= \int_{\mathbb{R}^d} \Big(\frac{1}{\sqrt{n}} \sum_{k=1}^n \sin(t'(X_k^* - \widehat{\mu}_n)) - \Im[c^*(t)] - c_{\mu_0}(t)\, t'\, l(X_k^*, \widehat{\mu}_n) \Big)^2 w(t)\, dt \\
&\quad + o_{P^*}(1).
\end{aligned}$$

Therefore, it suffices to show that $S_n^* \xrightarrow{d} Z_S + \mathbb{E}h_{\mu_0}(X_1, X_1)$, in probability, where

$$S_n^* := \int_{\mathbb{R}^d} \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n \sin(t'(X_k^* - \widehat{\mu}_n)) - \Im[c^*(t)] - c_{\mu_0}(t)\, t'\, l(X_k^*, \widehat{\mu}_n) \right)^2 w(t)\, dt.$$

In order to apply our results from Chapter 3, the validity of the assumptions (A1*), (A2*), and (A3*) has to be verified. The conditions (A1*), (A2*)(i), and (A3*)(ii) hold according to (S1)(ii) and (S2)(i). The moment constraints of (A3*)(i) and (A2*)(iii) follow from the fact that $\mathbb{E}^*\|l(X_1^*, \widehat{\mu}_n)\|_1^{2\bar{\nu}} = O_P(1)$ in conjunction with the distributional convergence of the bootstrap variables imply $\mathbb{E}^*\|l(X_1^*, \widehat{\mu}_n)\|_1^{2\nu} \xrightarrow{P} \mathbb{E}\|l(X_1, \mu_0)\|_1^{2\nu}$ for $\nu \in ((2-\delta)/(1-\delta), \bar{\nu})$. Moreover, the statistic S_n^* degenerates under the distribution of X_1^*. However, the direct application of Theorem 4.1 is not possible since S_n^* is not the bootstrap counterpart of S_n, i.e. the kernel functions of both statistics do not coincide. Still, Proposition 4.1 implies that $n\bar{V}_n^* \xrightarrow{d} Z_S + \mathbb{E}h_{\mu_0}(X_1, X_1)$, in probability. Here, the definition

$$\begin{aligned}
\bar{V}_n^* := \frac{1}{n^2} \sum_{j,k=1}^n \Big[& h(X_j^*, X_k^*, \widehat{\mu}_n) - \int_{\mathbb{R}^d} h(x, X_k^*, \widehat{\mu}_n)\, P_{X_1^*|X_1,\ldots,X_n}(dx) \\
& - \int_{\mathbb{R}^d} h(X_j^*, y, \widehat{\mu}_n)\, P_{X_1^*|X_1,\ldots,X_n}(dy) \\
& + \iint_{\mathbb{R}^d \times \mathbb{R}^d} h(x, y, \widehat{\mu}_n)\, P_{X_1^*|X_1,\ldots,X_n}(dx)\, P_{X_1^*|X_1,\ldots,X_n}(dy) \Big]
\end{aligned}$$

with

$$\begin{aligned}
h(x, y, \widehat{\mu}_n) = \int_{\mathbb{R}^d} & \big[\sin(t'(x - \widehat{\mu}_n)) - c_{\mu_0}(t)\, t'\, l(x, \widehat{\mu}_n) \big] \\
& \times \big[\sin(t'(y - \widehat{\mu}_n)) - c_{\mu_0}(t)\, t'\, l(y, \widehat{\mu}_n) \big]\; w(t)\, dt
\end{aligned}$$

is used. Actually, $n\,\bar{V}_n^*$ and S_n^* coincide which finally yields the assertion. □

The second result of Proposition 5.4 implies that the bootstrap-based test has asymptotically the correct size under the null hypothesis, i.e. $P(\widehat{S}_n > t^*_\alpha) \longrightarrow_{n\to\infty} \alpha$. Moreover, consistency and asymptotic unbiasedness against local alternatives of the form $\mathcal{H}_{1,n}$ of the bootstrap-aided test can be verified. Of course, non-trivial power against these Pitman alternatives would be desirable. Although Proposition 5.3 merely yields unbiasedness, we conjecture that our test has non-trivial power against certain alternatives of the structure $\mathcal{H}_{1,n}$. In the i.i.d. case for instance, one can verify non-trivial power for certain local alternatives based on the asymptotic results on U-statistics under local alternatives of Gregory [69] if the eigenfunctions of the integral equation (2.11) are known. However, since they are unknown in general and since things become even more involved in the case of dependent observations, we do not pursue these investigations here.

Corollary 5.1. *Suppose that the assumptions (S1) and (S2) are fulfilled and let $\alpha \in (0,1)$. The proposed bootstrap test based on the algorithm above satisfies*

$$\lim_{n\to\infty} P(\widehat{S}_n > t^*_\alpha) = \begin{cases} \alpha & \text{if } \mathcal{H}_0 \text{ is true and } \operatorname{var}(Z_S) > 0, \\ 1 & \text{if } \mathcal{H}_1 \text{ is true} \end{cases}$$

and under $\mathcal{H}_{1,n}$ and the additional assumptions of Proposition 5.3,

$$\liminf_{n\to\infty} P(\widehat{S}_{n,n} > t^*_\alpha) \geq \alpha.$$

Proof. Step 1: $\lim_{n\to\infty} P(\widehat{S}_n > t^*_\alpha) = \alpha$.
This is a consequence of Proposition 5.4.
Step 2: $\lim_{n\to\infty} P(\widehat{S}_n > t^*_\alpha) = 1$.
Note that $\widehat{S}^*_n - \bar{S}^*_n = o_{P^*}(1)$ holds under $\mathcal{H}_1$. The statistic $\bar{S}^*_n$ is the bootstrap counterpart of a degenerate V-statistic $\bar{S}_n$ with kernel

$$\begin{aligned} \bar{h}_{\mu_0}(x,y) := \int_{\mathbb{R}^d} \Big(&\big[\sin(t'(x-\mu_0)) - \Im[c_{\mu_0}(t)] - \Re[c_{\mu_0}(t)]\, t'\, l(x,\mu_0)\big] \\ &\times \big[\sin(t'(y-\mu_0)) - \Im[c_{\mu_0}(t)] - \Re[c_{\mu_0}(t)]\, t'\, l(y,\mu_0)\big] \Big)\, w(t)\, dt, \end{aligned}$$

where c_{μ_0} denotes the characteristic function of $X_1 - \mu_0$. In view of Lemma 3.7, the asymptotic (1-α)-quantiles, $\alpha \in (0,1)$, of $\bar{S}_n$ are bounded. One obtains equivalence of the limits of $\bar{S}_n$ and $\bar{S}^*_n$ in analogy to Proposition 5.4. Hence, the bootstrap quantiles t^*_α, $\alpha \in (0,1)$, are bounded in probability. Consequently, the claim follows immediately from Lemma 5.9.
Step 3: $\liminf_{n\to\infty} P(\widehat{S}_{n,n} > t^*_\alpha) \geq \alpha$.
First note that the bootstrap algorithm imitates the null situation. In conjunction with Proposition 5.4 and the continuity of the distribution function of Z_S, we obtain

$$P(t^*_\alpha \in [t_{\alpha+\delta}, t_{\alpha-\delta}]) \underset{n\to\infty}{\longrightarrow} 1 \quad \forall\, \delta > 0 \text{ with } \alpha \pm \delta \in (0,1), \tag{5.22}$$

where t_x denotes the $(1-x)$-quantile of the distribution of $Z_S + \mathbb{E}h_{\mu_0}(X_1, X_1)$. Let $\varepsilon > 0$ arbitrary but fixed, then we are to show $\liminf_{n\to\infty} P(\widehat{S}_{n,n} > t_\alpha^*) - \alpha \geq -\varepsilon$. According to step 1 of the proof it suffices to show that $\liminf_{n\to\infty}[P(\widehat{S}_{n,n} > t_\alpha^*) - P(\widehat{S}_n > t_\alpha^*)] > -\varepsilon$. The application of (5.22) and Proposition 5.3 lead to

$$
\begin{aligned}
&\liminf_{n\to\infty} \left[P(\widehat{S}_{n,n} > t_\alpha^*) - P(\widehat{S}_n > t_\alpha^*)\right] \\
&\geq \lim_{n\to\infty} \left[P(\widehat{S}_{n,n} > t_{\alpha-\delta}) - P(\widehat{S}_n > t_{\alpha+\delta})\right] \\
&\geq \lim_{n\to\infty} \left[P(\widehat{S}_n > t_{\alpha-\delta}) - P(\widehat{S}_n > t_{\alpha+\delta})\right] \\
&= P(Z_S + \mathbb{E}h_{\mu_0}(X_1, X_1) > t_{\alpha-\delta}) - P(Z_S + \mathbb{E}h_{\mu_0}(X_1, X_1) > t_{\alpha+\delta})
\end{aligned}
$$

which is less than $-\varepsilon$ for any sufficiently small $\delta > 0$ as the cumulative distribution function of Z_S is continuous. □

5.4 A goodness-of-fit test for the marginal distribution of a time series

5.4.1 Motivation

There is a great variety of tests concerning the problem whether the distribution of a sample belongs to a parametric class of distributions if the underlying observations are independent and identically distributed. The most popular tests are probably the Kolmogorov-Smirnov test, the Cramér-von Mises test, and the Anderson-Darling test.

However, in the case of dependent data the number of consistent tests is limited. Recently, Ignaccolo [75] as well as Munk et al. [91] developed results for α-mixing random variables generalizing Neyman's smooth test for a simple null hypothesis. Tests for the composite hypothesis based on the L_2-difference between a smoothed version of the parametric density estimate and a nonparametric estimator were considered by Neumann and Paparoditis [94] as well as by Fan and Ullah [66]. In both papers the underlying variables were assumed to arise from a β-mixing process. The bandwidths involved in the estimators were supposed to be asymptotically vanishing. A drawback resulting from the latter assumption is the loss of power against so-called Pitman local alternatives compared to approaches with fixed bandwidths; see Ghosh and Huang [68].

We consider a consistent test with fixed bandwidth for the composite hypothesis here. Our approach is based on the L_2-distance between the empirical characteristic function and the characteristic function with estimated parameters. These procedures are eminently suitable when the considered parametric family is characterized by a probability density function of complicated form but has a simple characteristic function. Examples include the variance gamma distribution and the NIG distribution. In the time series context, these distributions are widely applied in finance and turbulence, cf. Barndorff-Nielsen [8], Jensen and Lunde [76] or Rachev [97].

5.4.2 The test statistic and its asymptotic behaviour

We extend the test originally proposed by Fan [62] for i.i.d. random variables to weakly dependent observations. Suppose $X_1, \ldots, X_n$ are $\mathbb{R}^d$-valued observations with distribution P_X and consider the following test problem:

$$\mathcal{H}_0: \quad P_X \in \left\{P^\theta | \theta \in \Theta \subseteq \mathbb{R}^p\right\} \quad \text{vs.} \quad \mathcal{H}_1: \quad P_X \notin \left\{P^\theta | \theta \in \Theta \subseteq \mathbb{R}^p\right\}$$

which is equivalent to

$$\mathcal{H}_0: \quad c = c(\cdot, \theta_0) \quad \text{for some } \theta_0 \in \Theta \quad \text{vs.} \quad \mathcal{H}_1: \quad c \neq c(\cdot, \theta) \quad \forall\, \theta \in \Theta.$$

Here, c and $c(\cdot, \theta)$ denote the characteristic functions associated with P_X and $P^\theta, \theta \in \Theta$, respectively. Fan [62] suggested a test statistic of the form

$$\widehat{G}_n := n \int_{\mathbb{R}^d} \left| c_n(t) - c(t, \widehat{\theta}_n) \right|^2 w(t)\, dt,$$

where $\widehat{\theta}_n$ is a $\sqrt{n}$-consistent estimator of θ_0 and w denotes an appropriate weight function. In order to derive the limit distribution of $\widehat{G}_n$, we make the following assumptions:

(G1) **(i)** $(X_n)_{n\in\mathbb{N}}$ is a stationary τ-dependent process with $\mathbb{E}\|X_1\|_1^{(2-\delta)/(1-\delta)} < \infty$ and $\sum_{r=1}^\infty r(\tau_r)^{\delta^2} < \infty$ for some $\delta \in (0, 1)$.

(ii) The sequence of estimators $\widehat{\theta}_n$ admits the expansion

$$\widehat{\theta}_n - \theta_0 = \frac{1}{n} \sum_{k=1}^n l(X_k, \theta_0) + o_P(n^{-1/2}), \tag{5.23}$$

where $\mathbb{E} l(X_1, \theta_0) = 0$ and $\mathbb{E}\|l(X_1, \theta_0)\|_1^{2\nu} < \infty$ for some $\nu > (2-\delta)/(1-\delta)$. Moreover, $\|l(x, \theta_0) - l(\bar{x}, \theta_0)\|_1 \leq f_l(x, \bar{x}, \theta_0)\|x - \bar{x}\|_1$, where $f_l(\cdot, \theta_0)$ is symmetric and continuous and for any independent copy $\widetilde{X}_1$ of X_1 and some $A > 0$,

$$\begin{aligned} &\sup_{j,k\in\mathbb{N}} \mathbb{E} \max_{a\in[-A,A]^d} |f_l(X_j + a, X_k, \theta_0)|^{2(2-\delta)/(1-\delta)} \\ &+ \mathbb{E} \max_{a\in[-A,A]^d} |f_l(X_1 + a, \widetilde{X}_1, \theta_0)|^{2(2-\delta)/(1-\delta)} < \infty. \end{aligned}$$

(iii) The weight function $w: \mathbb{R}^d \to \mathbb{R}$ is measurable and positive a.e. w.r.t. the Lebesgue measure on $\mathbb{R}^d$. Additionally, it satisfies $\int_{\mathbb{R}^d} (1 + \|t\|_1^2)\, w(t)\, dt < \infty$.

(iv) The function $c(t, \cdot)$ is twice continuously differentiable for all $t \in \mathbb{R}^d$ with $\int_{\mathbb{R}^d} (\|t\|_1 \|c^{(1)}(t, \theta_0)\|_1 + \|c^{(1)}(t, \theta_0)\|_1^2)\, w(t)\, dt < \infty$. Additionally, there exists a neighbourhood $U(\theta_0) \subseteq \Theta$ such that for all $\theta \in U(\theta_0)$, every element of $c^{(2)}(\cdot, \theta)$ can be bounded by some function M with $\int_{\mathbb{R}^d} M^2(t)\, w(t)\, dt < \infty$.

Under the null hypothesis, the test statistic $\widehat{G}_n$ can be approximated by a degenerate V-statistic and thus the limit distribution can be derived similarly as in the previous section.

Theorem 5.2. *Suppose that the assumption (G1) holds. Then, under $\mathcal{H}_0$,*

(i) $\widehat{G}_n - G_n = o_P(1)$, *where*

$$G_n := n \int_{\mathbb{R}^d} \left| c_n(t) - c(t,\theta_0) - \frac{1}{n}\sum_{j=1}^{n} [l(X_j,\theta_0)]'\, c^{(1)}(t,\theta_0) \right|^2 w(t)\, dt,$$

(ii) moreover,

$$\widehat{G}_n \xrightarrow{d} Z_G + \mathbb{E}h_{\theta_0}(X_1,X_1).$$

Here, the distribution of $Z_G + \mathbb{E}h_{\theta_0}(X_1,X_1)$ *is the limit distribution of* $n\,V_n$ *defined in Theorem 3.2, where* V_n *denotes a* V*-statistic based on the underlying sample and whose kernel function is given by*

$$h_{\theta_0}(x,y) := \int_{\mathbb{R}^d} \left(\Re[g(x,t,\theta_0)]\, \Re[g(y,t,\theta_0)] + \Im[g(x,t,\theta_0)]\, \Im[g(y,t,\theta_0)]\right) w(t)\, dt.$$

Here, the function $g : \mathbb{R}^d \times \mathbb{R}^d \times \Theta \to \mathbb{C}$ *is defined by* $g(x,t,\theta) := e^{it'x} - c(t,\theta) - [l(x,\theta)]'\, c^{(1)}(t,\theta)$.

Proof. (i) Define $\widetilde{G}_n := n \int_{\mathbb{R}^d} \left| c_n(t) - c(t,\theta_0) - (\widehat{\theta}_n - \theta_0)'\, c^{(1)}(t,\theta_0) \right|^2 w(t)\, dt$.

Step 1: $\widehat{G}_n = \widetilde{G}_n + o_P(1)$.

Recall that for any characteristic function c, the equality $\overline{c(t)} = c(-t)$ holds for all $t \in \mathbb{R}^d$, where the bar denotes the complex conjugate. Taylor expansion gives

$$\begin{aligned}
\widehat{G}_n - \widetilde{G}_n = &- \frac{n}{2} \int_{\mathbb{R}^d} [c_n(t) - c(t,\theta_0)] \left(\widehat{\theta}_n - \theta_0\right)' c^{(2)}(-t,\widetilde{\theta}) \left(\widehat{\theta}_n - \theta_0\right) w(t)\, dt \\
&- \frac{n}{2} \int_{\mathbb{R}^d} [c_n(-t) - c(-t,\theta_0)] \left(\widehat{\theta}_n - \theta_0\right)' c^{(2)}(t,\widetilde{\theta}) \left(\widehat{\theta}_n - \theta_0\right) w(t)\, dt \\
&+ \frac{n}{2} \int_{\mathbb{R}^d} \left(\widehat{\theta}_n - \theta_0\right)' c^{(1)}(t,\theta_0) \left(\widehat{\theta}_n - \theta_0\right)' c^{(2)}(-t,\widetilde{\theta}) \left(\widehat{\theta}_n - \theta_0\right) w(t)\, dt \\
&+ \frac{n}{2} \int_{\mathbb{R}^d} \left(\widehat{\theta}_n - \theta_0\right)' c^{(1)}(-t,\theta_0) \left(\widehat{\theta}_n - \theta_0\right)' c^{(2)}(t,\widetilde{\theta}) \left(\widehat{\theta}_n - \theta_0\right) w(t)\, dt \\
&+ \frac{n}{4} \int_{\mathbb{R}^d} \left| \left(\widehat{\theta}_n - \theta_0\right)' c^{(2)}(t,\widetilde{\theta}) \left(\widehat{\theta}_n - \theta_0\right) \right|^2 w(t)\, dt
\end{aligned}$$

with some random $\widetilde{\theta}$ between $\widehat{\theta}_n$ and θ_0. According to Lemma 5.8 and the expansion (5.23), $\widehat{\theta}_n$ is a $\sqrt{n}$-consistent estimator. Therefore, the latter three summands vanish asymptotically due to the assumption (G1). The analysis of the remaining expressions are equal to each other. The absolute value of the first term can be bounded from above by

$$o_P(1) + O_P(1) \left(\int_{\mathbb{R}^d} |c_n(t) - c(t,\theta_0)|^2 w(t)\, dt \right)^{1/2} \left(\int_{\mathbb{R}^d} M^2(-t)\, w(t)\, dt \right)^{1/2}.$$

Similar to the consideration of the quantity T_1 in the proof of Lemma 5.7, asymptotic negligibility of the latter expression is a consequence of

$$\begin{aligned}\mathbb{E}|c_n(t)-c(t,\theta_0)|^2 &\leq \frac{2}{n}+\frac{2}{n^2}\sum_{j<k}\mathbb{E}\big[\sin(t'X_j)-\Im c(t,\theta_0)\big]\big[\sin(t'X_k)-\sin(t'\widetilde{X}_k)\big]\\ &\quad+\frac{2}{n^2}\sum_{j<k}\mathbb{E}\big[\cos(t'X_j)-\Re c(t,\theta_0)\big]\big[\cos(t'X_k)-\cos(t'\widetilde{X}_k)\big]\\ &\leq \frac{2}{n}+\frac{C\|t\|_1}{n}\sum_{r=1}^{\infty}\tau_r.\end{aligned} \tag{5.24}$$

Here, $\widetilde{X}_k$ denotes a suitable copy of X_k which is independent of X_j.
Step 2: $\widetilde{G}_n = G_n + o_P(1)$.
We split up as follows:

$$\begin{aligned}&\widetilde{G}_n - G_n =\\ &-n\int_{\mathbb{R}^d}[c_n(t)-c(t,\theta_0)]\Big(\widehat{\theta}_n-\theta_0-\frac{1}{n}\sum_{k=1}^{n}l(X_k,\theta_0)\Big)'c^{(1)}(-t,\theta_0)\,w(t)\,dt\\ &-n\int_{\mathbb{R}^d}[c_n(-t)-c(-t,\theta_0)]\Big(\widehat{\theta}_n-\theta_0-\frac{1}{n}\sum_{k=1}^{n}l(X_k,\theta_0)\Big)'c^{(1)}(t,\theta_0)\,w(t)\,dt\\ &+n\int_{\mathbb{R}^d}(\widehat{\theta}_n-\theta_0)'\,c^{(1)}(t,\theta_0)\,\Big[c^{(1)}(-t,\theta_0)\Big]'\Big(\widehat{\theta}_n-\theta_0-\frac{1}{n}\sum_{k=1}^{n}l(X_k,\theta_0)\Big)\,w(t)\,dt\\ &+n\int_{\mathbb{R}^d}\Big(\widehat{\theta}_n-\theta_0-\frac{1}{n}\sum_{k=1}^{n}l(X_k,\theta_0)\Big)'\\ &\qquad\times c^{(1)}(t,\theta_0)\,\Big[c^{(1)}(-t,\theta_0)\Big]'\Big(\frac{1}{n}\sum_{k=1}^{n}l(X_k,\theta_0)\Big)\,w(t)\,dt.\end{aligned}$$

According to inequality (5.24) it can be shown easily that the first two summands vanish asymptotically. The middle term is asymptotically negligible due to condition (G1)(ii). Applying Lemma 5.8 to the expression $n^{-1/2}\sum_{k=1}^{n}l(X_k,\theta_0)$, the third summand tends to zero as well.

(ii) The statistic G_n is of degenerate V-type with symmetric kernel. Under the assumption (G1), the kernel function satisfies the moment constraint

$$\sup_{k\in\mathbb{N}}\mathbb{E}|h_{\theta_0}(X_1,X_{1+k})|^{\nu}+\mathbb{E}|h_{\theta_0}(X_1,\widetilde{X}_1)|^{\nu}<\infty,$$

where $\widetilde{X}_1$ is an independent copy of X_1. Furthermore, note that h exhibits the continuity property

$$\begin{aligned}
|h_{\theta_0}(x,y) - h_{\theta_0}(\bar{x},\bar{y})| \le{}& C(\theta_0)\big\{(1+|f_l(x,\bar{x},\theta_0)|)(1+\|l(y,\theta_0)\|_1+\|l(\bar{y},\theta_0)\|_1) \\
&+ (1+|f_l(y,\bar{y},\theta_0)|)(1+\|l(x,\theta_0)\|_1+\|l(\bar{x},\theta_0)\|_1)\big\} \\
&\times [\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1] \\
=:{}& f_{\theta_0}(x,\bar{x},y,\bar{y})\,[\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1]\,.
\end{aligned} \tag{5.25}$$

In view of the assumption (G1), the function f_{θ_0} satisfies the condition (A4). Summing up, assertion (ii) is an immediate consequence of part (i) and Theorem 3.2.

□

Analogously to the test statistic of the previous section, $\widehat{G}_n$ is asymptotically unbounded if the alternative hypothesis $\mathcal{H}_1$ is true.

Lemma 5.10. *Suppose that (G1)(i) and (iii) hold. Moreover, assume that there exists a constant $\theta_0 \in \Theta$ such that $\widehat{\theta}_n \xrightarrow{P} \theta_0$. Then, under $\mathcal{H}_1$,*

$$P\left(\widehat{G}_n > K\right) \underset{n\to\infty}{\longrightarrow} 1, \quad \forall\, K < \infty.$$

Proof. It is sufficient to verify the existence of some $\eta > 0$ with $P(n^{-1}\widehat{G}_n > \eta) \longrightarrow_{n\to\infty} 1$. According to uniform continuity of characteristic functions, there exist $-\infty < T_{1,i} < T_{2,i} < \infty,\ i = 1,\dots,d$, such that

$$|c(t) - c(t,\theta_0)| > c_0, \forall\, t \in \Omega_T := [T_{1,1}, T_{2,1}] \times \cdots \times [T_{1,d}, T_{2,d}]$$

for some $c_0 > 0$. This implies

$$\begin{aligned}
&P\big(n^{-1}\widehat{G}_n > \eta\big) \\
&\ge P\Big(\int_{\Omega_T} \big|c_n(t) - c(t) + c(t) - c(t,\theta_0) - (\widehat{\theta}_n - \theta_0)'\, c^{(1)}(t,\theta_0) \\
&\qquad - \frac{1}{2}(\widehat{\theta}_n - \theta_0)'\, c^{(2)}(t,\widetilde{\theta})\,(\widehat{\theta}_n - \theta_0)\big|^2\, w(t)\, dt > \eta\Big) \\
&\ge P\Big(\int_{\Omega_T} \Big\{\big|c(t) - c(t,\theta_0)\big|^2 - 2\big|c(-t) - c(-t,\theta_0)\big|\big|c_n(t) - c(t) \\
&\qquad - (\widehat{\theta}_n - \theta_0)'\, c^{(1)}(t,\theta_0) - \frac{1}{2}(\widehat{\theta}_n - \theta_0)'\, c^{(2)}(t,\widetilde{\theta})\,(\widehat{\theta}_n - \theta_0)\big|\Big\}\, w(t)\, dt > \eta\Big) \\
&\ge P\Big(\int_{\Omega_T} \big|c(t) - c(t,\theta_0)\big|^2\, w(t)\, dt > 2\eta,\ 4\int_{\Omega_T} \Big\{\big|c_n(t) - c(t)\big| \\
&\qquad + \big|(\widehat{\theta}_n - \theta_0)'\, c^{(1)}(t,\theta_0) + \frac{1}{2}(\widehat{\theta}_n - \theta_0)' c^{(2)}(t,\widetilde{\theta})\,(\widehat{\theta}_n - \theta_0)\big|\Big\}\, w(t)\, dt \le \eta\Big)
\end{aligned}$$

for some random $\widetilde{\theta}$ between $\widehat{\theta}_n$ and θ_0. For sufficiently small $\eta > 0$ we obtain

$$\begin{aligned}
&P(n^{-1}\widehat{G}_n > \eta) \\
&\geq P\Big(\int_{\Omega_T} \Big\{\big|c_n(t) - c(t)\big| \\
&\qquad + \Big|(\widehat{\theta}_n - \theta_0)'\, c^{(1)}(t,\theta_0) + \frac{1}{2}(\widehat{\theta}_n - \theta_0)' c^{(2)}(t,\widetilde{\theta})\,(\widehat{\theta}_n - \theta_0)\Big|\Big\}\, w(t)\, dt \leq \eta/4\Big) \\
&\geq\; P\Big(\int_{\Omega_T} \Big|(\widehat{\theta}_n - \theta_0)'\, c^{(1)}(t,\theta_0) + \frac{1}{2}(\widehat{\theta}_n - \theta_0)'\, c^{(2)}(t,\widetilde{\theta})\,(\widehat{\theta}_n - \theta_0)\big|w(t)\, dt \leq \eta/8\Big) \\
&\quad + P\left(\int_{\Omega_T} \big|c_n(t) - c(t)\big|\, w(t)\, dt \leq \eta/8\right) - 1.
\end{aligned}$$

The first probability on the r.h.s. tends to one. Thus, it suffices to show that

$$P\left(\int_{\Omega_T} |c_n(t) - c(t)|\, w(t)\, dt < \eta/8\right) \underset{n\to\infty}{\longrightarrow} 1. \tag{5.26}$$

To this end, one can proceed as in the proof of Lemma 5.9. First, we consider

$$\lim_{\delta\to 0} \limsup_{n\to\infty} P\left(\sup_{\|s-t\|_1 \leq \delta} |c_n(s) - c(s) - c_n(t) + c(t)| \geq \varepsilon\right).$$

The expression tends to zero according to uniform Lipschitz continuity of $(c_n)_{n\in\mathbb{N}}$ and uniform continuity of c. It therefore remains to prove pointwise convergence of $c_n(t) \xrightarrow{P} c(t),\ t \in \mathbb{R}^d$ in order to deduce (5.26). This in turn follows from inequality (5.24) in the proof of Theorem 5.2. □

Finally, a result regarding the behaviour of the test statistic under local alternatives of the form

$$\mathcal{H}_{1,n} : \frac{dP_{X_{n,1}}}{dP_{X_1}} = 1 + \frac{g_n}{\sqrt{n}}, \text{ where } \|g_n - g\|_\infty \underset{n\to\infty}{\longrightarrow} 0 \text{ for some measurable bounded}$$
$$\text{function } g \text{ with } \int_{\mathbb{R}^d} e^{it'x} g(x) P_{X_1}(dx) \neq 0, \text{ for some } t \in \mathbb{R}^d,$$

is established.

Proposition 5.5. *Suppose that the conditions (A5) and (G1) are satisfied. Additionally, assume that* $\widehat{\theta}_{n,n} = \widehat{\theta}_n(X_{n,1}, \ldots, X_{n,n}) = n^{-1}\sum_{k=1}^n l(X_{n,k}, \theta_0) + o_P(n^{-1/2})$ *with*

$$\sup_{\substack{1\leq j,k\leq n,\\ n\in\mathbb{N}}} \mathbb{E} \max_{a\in[-A,A]^d} |f_l(X_{n,j} + a, X_{n,k})|^{2(2-\delta)/(1-\delta)} < \infty.$$

Then, as $n \to \infty$,

$$\widehat{G}_{n,n} \xrightarrow{d} Z_{G,loc} + \mathbb{E}h_{\theta_0}(X_1, X_1).$$

Here, $\widehat{G}_{n,n} := n\int_{\mathbb{R}^d} \big|c_{n,n}(t) - c(t, \widehat{\theta}_{n,n})\big|^2 w(t)\, dt$ *with* $c_{n,n}(t) := n^{-1}\sum_{k=1}^n e^{it'X_{n,k}}$. *The distribution of* $Z_{G,loc} + \mathbb{E}h_{\theta_0}(X_1, X_1)$ *is the limit distribution of* $n\, V_{n,n}$ *defined in Proposition 5.1, where* $V_{n,n}$ *denotes a* V*-statistic based on the underlying sample and whose kernel function is given by* h_{θ_0}. *If additionally* $\mathrm{var}(Z_G) > 0$, *then*

$$\liminf_{n\to\infty} \Big[P(\widehat{G}_{n,n} > x) - P(\widehat{G}_n > x)\Big] \geq 0, \quad \forall\, x \in \mathbb{R}.$$

Proof. We simply imitate the proof of Proposition 5.3 and assume w.l.o.g. that $\|g_n\|_\infty < \infty,\ \forall\, n \in \mathbb{N}$.

Step 1: $n\, \widehat{G}_{n,n} \xrightarrow{d} Z_{G,loc} + \mathbb{E} h_{\theta_0}(X_1, X_1)$.

To this end, the statistic

$$G_{n,n} := n \int_{\mathbb{R}^d} \left| c_{n,n}(t) - c(t,\theta_0) - n^{-1} \sum_{k=1}^{n} [l(X_{n,k},\theta_0)]'\, c^{(1)}(t,\theta_0) \right| w(t)\, dt$$

is introduced. It can be proved that $\widehat{G}_{n,n} - G_{n,n} = o_P(1)$ by carrying out slight modifications in the verification of Theorem 5.2(i). The definition of

$$\widetilde{G}_{n,n} := n \int_{\mathbb{R}^d} \left| c_{n,n}(t) - c(t,\theta_0) - (\widehat{\theta}_{n,n} - \theta_0)' c^{(1)}(t,\theta_0) \right|^2 w(t)\, dt$$

leads to $\widehat{G}_{n,n} - \widetilde{G}_{n,n} = o_P(1)$ in complete accordance to step 1 of the proof of Theorem 5.2 if $n\, \mathbb{E}\, |c_{n,n}(t) - c(t,\theta_0)|^2 = O(1)\, \|t\|_1$. This in turn follows from Lemma 5.8 and the relation

$$\mathbb{E}\, |c_{n,n}(t) - c(t,\theta_0)|^2 \le 2\, \mathbb{E} \left| c_{n,n}(t) - n^{-1/2} \int_{\mathbb{R}^d} g_n(x) e^{it'x} P_{X_1}(dx) - c(t,\theta_0) \right|^2 + O(n^{-1}).$$

Note that that additionally $n^{-1/2} \sum_{k=1}^n l(X_{n,k},\theta_0) = O_P(1)$. Hence, we immediately obtain $\widetilde{G}_{n,n} - G_{n,n} = o_P(1)$.

Moreover, if the condition (A6) is satisfied, Proposition 5.1 can be invoked to deduce the limit distribution of $G_{n,n}$ and $\widehat{G}_{n,n}$, respectively. The assumption (A6)(i) trivially holds in the present context. Additionally,

$$\begin{aligned} &\sup_{\substack{1 \le k < n,\\ n \in \mathbb{N}}} \mathbb{E} |h_{\theta_0}(X_{n,1}, X_{n,1+k})|^\nu \\ &\le \sup_{\substack{1 \le k < n,\\ n \in \mathbb{N}}} C\, \mathbb{E} \Big[1 + \|l(X_{n,1},\theta_0)\|_1^\nu + (\|l(X_{n,1},\theta_0)\|_1 \|l(X_{n,1+k},\theta_0)\|_1)^\nu \Big] \end{aligned}$$

is finite, which implies (A6)(ii). According to the definition of f_{θ_0} in (5.25) we obtain (A6)(iii). Thus, Proposition 5.1 implies $\widehat{G}_{n,n} \xrightarrow{d} Z_{G,loc} + \mathbb{E} h_{\theta_0}(X_1, X_1)$.

Step 2: $\liminf_{n\to\infty} [P(\widehat{G}_{n,n} > x) - P(\widehat{G}_n > x)] \ge 0, \quad \forall\, x \in \mathbb{R}$.

After deriving an alternative representation of the corresponding limit distributions similarly as in the proof of Proposition 5.3, we can deduce the assertion from Lemma 5.3. The modified representation is obtained when we plug in an adjusted definition of $\bar{h}_c$ into the calculations of Step 2 of the proof of Proposition 5.3. In order to define the new version of the truncated kernel $\bar{h}_c$, we introduce functions h_i by $h_1(x,t) := \cos(t'x) - \Re c(t) - [l(x,\theta_0)]' \Re c^{(1)}(t)$ and $h_2(x,t) := \sin(t'x) - \Im c(t) - [l(x,\theta_0)]' \Im c^{(1)}(t)$, functions $g_{c,i}$ by $g_{c,i}(t) := \max_{x \in [-c,c]^d} |h_i(x,t)|$ and functions $h_i^{(c)}$ by

$$h_i^{(c)}(x,t) := \begin{cases} h_i(x,t) & \text{for } |h_i(x,t)| \le g_{c,i}(t), \\ -g_{c,i}(t) & \text{for } h_i(x,t) < -g_{c,i}(t), \\ g_{c,i}(t) & \text{for } h_i(x,t) > g_{c,i}(t) \end{cases}$$

for $i = 1, 2$. Now we are in the position to define suitable truncated versions of h_{θ_0}, i.e.

$$\begin{aligned}
&\bar{h}_c(x,y) := \\
&\int_{\mathbb{R}^d} \Big\{ \Big[h_1^{(c)}(x,t) - \int_{\mathbb{R}^d} h_1^{(c)}(z,t) P_{X_1}(dz) \Big] \Big[h_1^{(c)}(y,t) - \int_{\mathbb{R}^d} h_1^{(c)}(z,t) P_{X_1}(dz) \Big] \\
&+ \Big[h_2^{(c)}(x,t) - \int_{\mathbb{R}^d} h_2^{(c)}(z,t) P_{X_1}(dz) \Big] \Big[h_2^{(c)}(y,t) - \int_{\mathbb{R}^d} h_2^{(c)}(z,t) P_{X_1}(dz) \Big] \Big\} w(t)\, dt.
\end{aligned}$$

Under our assumptions we have

$$\big| \bar{h}_c(x,y) - \bar{h}_c(\bar{x},\bar{y}) \big| \le C f_{\theta_0}(x,\bar{x},y,\bar{y}) \, [\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1]$$

and, additionally, the conditions 1.,2., and 4. of Remark 3.3 are satisfied. The assertion follows with the same arguments as they were used to prove Proposition 5.3 since now

$$\begin{aligned}
z' \bar{\Delta}_c^{(K,L)} z = \int_{\mathbb{R}^d} &\left(\sum_{k=1}^{(2K+1)^d} z_k \int_{\mathbb{R}^d} \Phi_{J(L),k} [h_1^{(c)}(x,t) - \mathbb{E} h_1^{(c)}(X_1,t)] \, dx \right)^2 \\
&+ \left(\sum_{k=1}^{(2K+1)^d} z_k \int_{\mathbb{R}^d} \Phi_{J(L),k} [h_2^{(c)}(x,t) - \mathbb{E} h_2^{(c)}(X_1,t)] \, dx \right)^2 w(t)\, dt \ge 0.
\end{aligned}$$

□

5.4.3 Bootstrapping critical values

The limit distribution of the test statistic $\widehat{G}_n$ under the null hypothesis has a complex structure and depends on the unknown parameter θ_0 in a complicated way. Therefore, difficulties occur as soon as asymptotic critical values of the test have to be determined. We propose to invoke a parametric bootstrap procedure to circumvent these problems. For this purpose we assume:

(G2) **(i)** The condition (A1*) holds with $\sum_{r=1}^{\infty} r\,(\bar{\tau}_r)^{\delta^2} < \infty$. Moreover, $\mathbb{E}^* \|X_1^*\|_1^{(2-\delta)/(1-\delta)} \xrightarrow{P} \mathbb{E}\|X_1\|_1^{(2-\delta)/(1-\delta)}$.

(ii) The sequence of estimators $\widehat{\theta}_n^*$ admits the expansion

$$\widehat{\theta}_n^* - \widehat{\theta}_n = \frac{1}{n} \sum_{k=1}^{n} l(X_k^*, \widehat{\theta}_n) + o_{P^*}(n^{-1/2}),$$

where $\mathbb{E}^* l(X_1^*, \widehat{\theta}_n) = 0$ and $\mathbb{E}^* \|l(X_1^*, \widehat{\theta}_n)\|_1^{2\nu} = O_P(1)$ for some $\nu > (2-\delta)/(1-\delta)$. Additionally, $\|l(x,\widehat{\theta}_n) - l(\bar{x},\widehat{\theta}_n)\|_1 \le f_l(x,\bar{x},\widehat{\theta}_n)\|x-\bar{x}\|_1$ with

$$P\left(\sup_{j,k\in\mathbb{N}} \mathbb{E}^* \max_{a\in[-A,A]^d} |f_l(X_j^*+a, X_k^*, \widehat{\theta}_n)|^{2(2-\delta)/(1-\delta)} \le K \right) \underset{n\to\infty}{\longrightarrow} 1$$

and $P\left(\mathbb{E}^* \max_{a\in[-A,A]^d} |f_l(X_1^*+a, \widetilde{X}_1^*, \widehat{\theta}_n)|^{2(2-\delta)/(1-\delta)} \le K \right) \longrightarrow_{n\to\infty} 1$ for a $K < \infty$ and some $A > 0$, where $\widetilde{X}_1^*$ denotes an independent copy of X_1^*, conditionally on $X_1, \dots, X_n$.

(iii) The first derivative of c w.r.t. θ satisfies the inequality $\int_{\mathbb{R}^d} \left(\|t\|_1 \, \|c^{(1)}(t,\theta)\|_1 + \|c^{(1)}(t,\theta)\|_1^2\right) w(t)\, dt < \infty$ for all θ in some neighbourhood $U(\theta_0) \subseteq \Theta$ of θ_0.

When bootstrapping the test statistic $\widehat{G}_n$, similar problems as in the previous section occur. Again the bootstrap counterpart of the approximating V-statistic G_n is not degenerate in general. Thus, in order to establish a consistent bootstrap method to determine critical values of the test statistic, the bootstrap statistic has to be degenerated artificially. For this purpose we suggest the application of the following bootstrap algorithm:

1. Determine $\widehat{\theta}_n$.
2. Generate $X_1^*, \ldots, X_n^*$ such that (G2)(i) holds.
3. Determine $\widehat{\theta}_n^*$ such that (G2)(ii) is satisfied.
4. Compute the bootstrap test statistic
$$\widehat{G}_n^* := n \int_{\mathbb{R}^d} \left| c_n^*(t) - c^*(t) + c(t, \widehat{\theta}_n) - c(t, \widehat{\theta}_n^*) \right|^2 w(t)\, dt.$$
Here, c^* is the characteristic function of X_1^*, conditionally on $X_1, \ldots, X_n$, and c_n^* denotes its empirical counterpart, i.e. $c_n^*(t) = n^{-1} \sum_{k=1}^n e^{it'X_k^*}$.
5. Define the critical value t_α^* as the $(1-\alpha)$-quantile of the (conditional) distribution of $\widehat{G}_n^*$. Reject $\mathcal{H}_0$ if $\widehat{G}_n > t_\alpha^*$.

Remark 5.5. The bootstrap characteristic function c^* is often unknown and has to be approximated by simulation. However, note that there are cases where the bootstrap variables can be generated such that $c^*(t) = c(t, \widehat{\theta}_n)$, see Taufer and Leonenko [106]. In these situations, $\widehat{G}_n^*$ actually is the bootstrap counterpart of $\widehat{G}_n$.

Proposition 5.6. *Suppose that the assumptions (G1) and (G2) are satisfied. Then, under* $\mathcal{H}_0$*,*
$$\widehat{G}_n^* \xrightarrow{d} Z_G + \mathbb{E} h_{\theta_0}(X_1, X_1) \quad \text{in probability,}$$
as $n \to \infty$*, where* Z_G *is defined as in Theorem 5.2. Moreover, if* $\operatorname{var}(Z_G) > 0$*,*
$$\sup_{-\infty < x < \infty} \left| P(\widehat{G}_n^* \le x | X_1, \ldots, X_n) - P(\widehat{G}_n \le x) \right| \xrightarrow{P} 0.$$

Proof. Similarly to step 1 and step 2 of the proof of Theorem 5.2, the bootstrap statistic can be approximated as follows:
$$\begin{aligned} \widehat{G}_n^* &= n \int_{\mathbb{R}^d} \left| c_n^*(t) - c^*(t) - (\widehat{\theta}_n^* - \widehat{\theta}_n)' c^{(1)}(t, \widehat{\theta}_n) \right|^2 w(t)\, dt + o_{P^*}(1) \\ &= n \int_{\mathbb{R}^d} \left| c_n^*(t) - c^*(t) - \Big(\frac{1}{n} \sum_{j=1}^n l(X_j^*, \widehat{\theta}_n) \Big)' c^{(1)}(t, \widehat{\theta}_n) \right|^2 w(t)\, dt + o_{P^*}(1). \end{aligned}$$

Thus, it suffices to show that $G_n^* \xrightarrow{d} Z_G + \mathbb{E}h_{\theta_0}(X_1, X_1)$, in probability, where

$$G_n^* := n \int_{\mathbb{R}^d} \Big| c_n^*(t) - c^*(t) - \Big(\frac{1}{n}\sum_{j=1}^n l(X_j^*, \widehat{\theta}_n)\Big)' c^{(1)}(t, \widehat{\theta}_n)\Big|^2 w(t)\, dt.$$

Under the assumptions (G1) and (G2), the prerequisites of Proposition 4.1 are fulfilled which implies that $n\,\bar{V}_n^* \xrightarrow{d} Z_G + \mathbb{E}h_{\theta_0}(X_1, X_1)$, in probability, with

$$\begin{aligned}\bar{V}_n^* =& \frac{1}{n^2}\sum_{j,k=1}^n \Big[h(X_j^*, X_k^*, \widehat{\theta}_n) - \int_{\mathbb{R}^d} h(x, X_k^*, \widehat{\theta}_n)\, P_{X_1^*|X_1,\dots,X_n}(dx)\\ &- \int_{\mathbb{R}^d} h(X_j^*, y, \widehat{\theta}_n)\, P_{X_1^*|X_1,\dots,X_n}(dy)\\ &+ \iint_{\mathbb{R}^d\times\mathbb{R}^d} h(x, y, \widehat{\theta}_n)\, P_{X_1^*|X_1,\dots,X_n}(dx)\, P_{X_1^*|X_1,\dots,X_n}(dy)\Big]\end{aligned}$$

and h is obtained via the substitution of θ_0 by $\widehat{\theta}_n$ in the definition of the kernel function of G_n. Actually, $n\,\bar{V}_n^*$ and G_n^* coincide:

$$\begin{aligned}\bar{n}\, V_n^* =& \frac{1}{n}\sum_{j,k=1}^n \int_{\mathbb{R}^d} \Big\{\Re\Big[e^{it'X_j^*} - c(t, \widehat{\theta}_n) - [l(X_j^*, \widehat{\theta}_n)]'\, c^{(1)}(t, \widehat{\theta}_n)\Big]\\ &\times \Re\Big[e^{it'X_k^*} - c(t, \widehat{\theta}_n) - [l(X_k^*, \widehat{\theta}_n)]'\, c^{(1)}(t, \widehat{\theta}_n)\Big]\\ &+ \Im\Big[e^{it'X_j^*} - c(t, \widehat{\theta}_n) - [l(X_j^*, \widehat{\theta}_n)]'\, c^{(1)}(t, \widehat{\theta}_n)\Big]\\ &\times \Im\Big[e^{it'X_k^*} - c(t, \widehat{\theta}_n) - [l(X_k^*, \widehat{\theta}_n)]'\, c^{(1)}(t, \widehat{\theta}_n)\Big]\\ &- \Re\Big[c^*(t) - c(t, \widehat{\theta}_n)\Big]\, \Re\Big[e^{it'X_k^*} - c(t, \widehat{\theta}_n) - [l(X_k^*, \widehat{\theta}_n)]'\, c^{(1)}(t, \widehat{\theta}_n)\Big]\\ &- \Re\Big[c^*(t) - c(t, \widehat{\theta}_n)\Big]\, \Re\Big[e^{it'X_j^*} - c(t, \widehat{\theta}_n) - [l(X_j^*, \widehat{\theta}_n)]'\, c^{(1)}(t, \widehat{\theta}_n)\Big]\\ &- \Im\Big[c^*(t) - c(t, \widehat{\theta}_n)\Big]\, \Im\Big[e^{it'X_k^*} - c(t, \widehat{\theta}_n) - [l(X_k^*, \widehat{\theta}_n)]'\, c^{(1)}(t, \widehat{\theta}_n)\Big]\\ &- \Im\Big[c^*(t) - c(t, \widehat{\theta}_n)\Big]\, \Im\Big[e^{it'X_j^*} - c(t, \widehat{\theta}_n) - [l(X_j^*, \widehat{\theta}_n)]'\, c^{(1)}(t, \widehat{\theta}_n)\Big]\\ &+ \Big|c^*(t) - c(t, \widehat{\theta}_n)\Big|^2\Big\} w(t)\, dt\\ =& \frac{1}{n}\sum_{j,k=1}^n \int_{\mathbb{R}^d} \Big\{\Re\Big[e^{it'X_j^*} - c(t, \widehat{\theta}_n) - l'(X_j^*, \widehat{\theta}_n) c^{(1)}(t, \widehat{\theta}_n)\Big]\\ &\times \Re\Big[e^{it'X_k^*} - c^*(t) - [l(X_k^*, \widehat{\theta}_n)]'\, c^{(1)}(t, \widehat{\theta}_n)\Big]\\ &+ \Im\Big[e^{it'X_j^*} - c(t, \widehat{\theta}_n) - [l(X_j^*, \widehat{\theta}_n)]'\, c^{(1)}(t, \widehat{\theta}_n)\Big]\\ &\times \Im\Big[e^{it'X_k^*} - c^*(t) - [l(X_k^*, \widehat{\theta}_n)]'\, c^{(1)}(t, \widehat{\theta}_n)\Big]\\ &- \Re\Big[c^*(t) - c(t, \widehat{\theta}_n)\Big]\, \Re\Big[e^{it'X_k^*} - c^*(t) - [l(X_k^*, \widehat{\theta}_n)]'\, c^{(1)}(t, \widehat{\theta}_n)\Big]\\ &- \Im\Big[c^*(t) - c(t, \widehat{\theta}_n)\Big]\, \Im\Big[e^{it'X_k^*} - c^*(t) - [l(X_k^*, \widehat{\theta}_n)]'\, c^{(1)}(t, \widehat{\theta}_n)\Big]\Big\} w(t)\, dt\\ =& G_n^*.\end{aligned}$$

Therefore, $G_n^* \xrightarrow{d} Z_G + \mathbb{E}h_{\theta_0}(X_1, X_1)$ in probability, which completes the proof. $\square$

The above proposition assures that, under $\mathcal{H}_0$, the bootstrap-based test is asymptotically of correct size. Moreover, our procedure is consistent and asymptotically unbiased under Pitman local alternatives. To sum up, we obtain the subsequent result.

Corollary 5.2. *Suppose that the assumptions (G1) and (G2) are fulfilled and let $\alpha \in (0,1)$. The proposed bootstrap test based on the algorithm above satisfies*

$$\lim_{n\to\infty} P(\widehat{G}_n > t_\alpha^*) = \begin{cases} \alpha & \textit{if } \mathcal{H}_0 \textit{ is true and } \operatorname{var}(Z_G) > 0, \\ 1 & \textit{if } \mathcal{H}_1 \textit{ is true.} \end{cases}$$

Moreover, under $\mathcal{H}_{1,n}$ and the additional assumptions of Proposition 5.5,

$$\liminf_{n\to\infty} P(\widehat{G}_{n,n} > t_\alpha^*) \geq \alpha,$$

that is, the test is asymptotically unbiased.

Proof. *Step 1:* $\lim_{n\to\infty} P(\widehat{G}_n > t_\alpha^*) = \alpha$.
This assertion follows from Proposition 5.6.
Step 2: $\lim_{n\to\infty} P(\widehat{G}_n > t_\alpha^*) = 1$.
Note that $\widehat{G}_n^* - G_n^* = o_{P^*}(1)$ holds under $\mathcal{H}_1$ as well and that G_n^* is still degenerate. Thus, the bootstrap procedure imitates the null situation $c(t) = c(t, \theta_0)$. Proposition 5.6 implies that the bootstrap quantiles are bounded in probability. Now, the claim follows immediately from Lemma 5.10.
Step 3: $\liminf_{n\to\infty} P(\widehat{G}_{n,n} > t_\alpha^*) \geq \alpha$.
With the same arguments as in the proof of Corollary 5.1, this inequality can be deduced from Proposition 5.5. $\square$

5.5 Consistent model specification tests based on parametric bootstrap

5.5.1 Survey of the literature and preliminaries

This part of the thesis is concerned with goodness-of-fit tests for parametric models of the conditional mean of a time series. More precisely, we consider a stationary time series $(Y_t)_t$ with values in $\mathbb{R}^d$ and derive a test for the problem whether the conditional mean of Y_t given some set of information I_t at time t belongs to a specific parametric family, i.e.

$$\begin{aligned} \mathcal{H}_0: \quad & P(\mathbb{E}(Y_t|I_t) = g(I_t, \theta_0)) &= 1 \quad & \text{for some } \theta_0 \in \Theta \subseteq \mathbb{R}^q \quad \text{vs.} \\ \mathcal{H}_1: \quad & P(\mathbb{E}(Y_t|I_t) = g(I_t, \theta)) &< 1 \quad & \forall\, \theta \in \Theta \subseteq \mathbb{R}^q. \end{aligned}$$

If Y_t is integrable, we obtain the tautological expression

$$Y_t = \mathbb{E}(Y_t|I_t) + \epsilon_t,$$

where the conditional expectation of $\epsilon_t = Y_t - \mathbb{E}(Y_t|I_t)$ given I_t is equal to zero almost surely.

In the context of time series, the information variable I_t may depend on lagged values of the response process $(Y_t)_t$. Both cases, I_t is finite-dimensional as well as I_t is infinite-dimensional, are treated in the literature. Tests for the latter case with $d = 1$ were derived for example by de Jong [46], Bierens and Ploberger [15] as well as by Escanciano [59]. We restrict ourselves to the finite-dimensional case here. Still, our method leads to tests that are consistent against a broad class of models such as linear and various nonlinear (auto-)regressions.

There are basically two approaches to establish consistent tests. Kernel-based tests with vanishing bandwidth were for instance considered by Fan and Li [63], Hjellvik, Yao and Tjøstheim [72], Fan and Li [64] as well as by Kreiss, Neumann and Yao [82]. While Fan and Li concentrated on the asymptotics under null and alternative hypotheses, the both other papers investigated a parametric bootstrap method and a wild bootstrap procedure, respectively.

The second approach extends the integrated conditional moment test of Bierens [14] towards dependent observations, cf. Koul and Stute [81], Escanciano [60] as well as Escanciano and Jacho-Chávez [61]. The first two papers are concerned with the behaviour of the respective test statistics under $\mathcal{H}_0$ and $\mathcal{H}_1$ for real-valued response variables. Additionally, Escanciano [60] investigated the asymptotics under Pitman local alternatives and justified a wild bootstrap method. Both articles rely on the asymptotic behaviour of residual marked empirical processes. Based on the work of Escanciano [60], Escanciano and Jacho-Chávez [61] developed a principal components decomposition of Cramér-von Mises types of tests. They approximated the corresponding critical values with the aid of Monte Carlo methods.

A comparative overview of both approaches is provided by Fan and Li [65]. In particular they investigated the behaviour under local alternatives. While the Bierens-type tests are more powerful against Pitman alternatives, the kernel-based method with vanishing bandwidth can detect alternatives characterized by sharp peaks with faster rate.

Here, kernel-based tests of L_2-type with fixed bandwidth are considered. Note that we allow for vector-valued response variables. Alternatively to Escanciano [60] who invoked empirical process theory in the case of real-valued response variables, we employ our results on degenerate U- and V-type statistics. Let $((Y_t', I_t')')_{t\in\mathbb{Z}}$ be a strictly stationary process, where the marginals of $(Y_t)_t$ have values in $\mathbb{R}^d$. The process $(Y_t)_t$ is assumed to be nonlinear autoregressive with exogenous terms (NARX), i.e.

$$Y_t := G(Y_{t-p}, \ldots, Y_{t-1}, Z_{t-\bar{p}+1}, \ldots, Z_{t-1}, Z_t) + \epsilon_t \tag{5.27}$$

and $I_t := (Y_{t-p}', \ldots, Y_{t-1}', Z_{t-\bar{p}+1}', \ldots, Z_{t-1}', Z_t')'$. Here, $((\epsilon_t', Z_t')')_t$ is a sequence of i.i.d. integrable $\mathbb{R}^{d+m}$-valued random variables with independent components ϵ_t and Z_t and $\mathbb{E}\epsilon_1 = 0_d$.

We consider test statistics of U- and V-type with estimated parameters $\widehat{\theta}_n$,

$$\widehat{T}_n^{(u)} := \frac{1}{n} \sum_{\substack{j,k=1\\ j\neq k}}^{n} [Y_j - g(I_j, \widehat{\theta}_n)]' \, K(I_j, I_k, \widehat{\theta}_n) \, [Y_k - g(I_k, \widehat{\theta}_n)],$$

$$\widehat{T}_n^{(v)} := \frac{1}{n} \sum_{j,k=1}^{n} [Y_j - g(I_j, \widehat{\theta}_n)]' \, K(I_j, I_k, \widehat{\theta}_n) \, [Y_k - g(I_k, \widehat{\theta}_n)]$$

with a diagonal matrix $K = diag(k_1, \ldots, k_d)$ of kernels $k_i : \mathbb{R}^{dp+m\bar{p}} \times \Theta \to \mathbb{R}$. Both statistics have been applied by Bartels [9] to independent observations.

The first statistic can be interpreted as a leave-one-out fixed-kernel estimator of $\mathbb{E}([Y_1 - g(I_1, \theta_0)]\mathbb{E}[Y_1 - g(I_1, \theta_0)|I_1] \, p(I_1))$ multiplied with the sample size, where p denotes the (unknown) density of I_1 and $d = 1$. For details, see Li and Wang [88]. Note that this quantity vanishes under the null hypothesis. Both statistics $\widehat{T}_n^{(u)}$ and $\widehat{T}_n^{(v)}$ only differ in view of the diagonal terms. A decision based on the U-type statistic does not take $n^{-1} \sum_{k=1}^{n} [Y_k - g(I_k, \widehat{\theta}_n)]' K(I_k, I_k, \widehat{\theta}_n)[Y_k - g(I_k, \widehat{\theta}_n)]$ into account. If, similar to the kernel-based tests with vanishing bandwidth, $k_i(x, y, \theta) = \bar{k}_i((x-y)/h, \theta)$ for some functions $\bar{k}_i$, $i = 1, \ldots, d$, and $h > 0$, this diagonal expression may be interpreted as an estimator of the weighted variances $\sum_{i=1}^{d} k_i(0, 0, \theta) \, \mathrm{var}(\epsilon_{1,i})$ of the components of the innovation vector. In general, this expression has no impact on the validity of $\mathcal{H}_0$.

Statistics of the form of $\widehat{T}_n^{(v)}$ include Bierens-type test statistics, i.e. $\int [\, n^{-1/2} \sum_{j=1}^{n} (Y_j - g(I_j, \widehat{\theta}_n)) \, w(X_j, t) \,]^2 \, \Psi(dt)$; $d = 1$. Here, Ψ denotes an integrating function and w is a weight function. The corresponding kernel of $\widehat{T}_n^{(v)}$ has the representation $K(x, y, \theta) = \int w(x, t) \, w(y, t) \, \Psi(dt)$. Eventually, note that we allow for parametric kernels. This might be of interest if one intends to direct the power towards special alternatives.

5.5.2 Asymptotic behaviour of the test statistics

In order to derive the asymptotic distributions of the test statistics $\widehat{T}_n^{(u)}$ and $\widehat{T}_n^{(v)}$ under the null hypothesis, certain smoothness and moment constraints on the conditional mean function $g : \mathbb{R}^{dp+m\bar{p}} \times \Theta \to \mathbb{R}$ are required:

(M1) **(i)** The function g satisfies $\mathbb{E}\|g(x, Z_1, \theta_0)\|_2 < \infty$ for some $x \in \mathbb{R}^{dp+m(\bar{p}-1)}$. Furthermore it admits the estimate

$$\begin{aligned} &\|g(y_1, \ldots, y_p, z_1, \ldots, z_{\bar{p}}, \theta) - g(\bar{y}_1, \ldots, \bar{y}_p, \bar{z}_1, \ldots, \bar{z}_{\bar{p}}, \theta)\|_2 \\ &\le \sum_{j=1}^{P} H_j(z_{\bar{p}}, \theta) \|(y_j', z_j')' - (\bar{y}_j', \bar{z}_j')'\|_2 \\ &\quad + H_{P+1}(\bar{y}_1, \ldots \bar{y}_p, \bar{z}_1, \ldots, \bar{z}_{\bar{p}-1}, \bar{z}_{\bar{p}}, z_{\bar{p}}, \theta) \|z_{\bar{p}} - \bar{z}_{\bar{p}}\|_2 \end{aligned} \tag{5.28}$$

with $\theta = \theta_0$, $P := \max\{p, \bar{p} - 1\}$, and continuous functions $(H_j(\cdot, \theta_0))_{j=1}^{P+1}$ such that $\sum_{j=1}^{P} \mathbb{E}|H_j(Z_1,\theta_0)| < 1$. Moreover, the moments

$$\mathbb{E} \max_{a\in[-A,A]^{dp+m\bar{p}}} |H(I_1+a, I_1, \theta_0)|^{4+\varepsilon},\ \mathbb{E} \max_{a\in[-A,A]^{dp+m\bar{p}}} |H(I_1, I_1+a, \theta_0)|^{4+\varepsilon},$$
$$\mathbb{E} \max_{a\in[-A,A]^{dp+m\bar{p}}} |H(I_1+a, \widetilde{I}_1, \theta_0)|^{4+\varepsilon},\ \mathbb{E} \max_{a\in[-A,A]^{dp+m\bar{p}}} |H(I_1, \widetilde{I}_1+a, \theta_0)|^{4+\varepsilon}$$

of the function $H := \sum_{j=1}^{P+1} H_j$ are finite for some $\varepsilon, A > 0$ and an independent copy $\widetilde{I}_1$ of I_1.

(ii) The function $g(x,\cdot)$ is three times continuously differentiable for all $x \in \mathbb{R}^{dp+m\bar{p}}$. The components of its first two partial derivatives w.r.t. θ satisfy $\sum_{i=1}^{d}\sum_{\alpha,\beta=1}^{q} \mathbb{E}\left[|g^{(1)}_{i;\alpha}(I_1,\theta_0)| + |g^{(2)}_{i;\alpha,\beta}(I_1,\theta_0)|\right]^{4+\varepsilon} < \infty$ and

$$\begin{aligned}&\sum_{i=1}^{d}\sum_{\alpha,\beta=1}^{q} \left[|g^{(1)}_{i;\alpha}(x,\theta) - g^{(1)}_{i;\alpha}(\bar{x},\theta)| + |g^{(2)}_{i;\alpha,\beta}(x,\theta) - g^{(2)}_{i;\alpha,\beta}(\bar{x},\theta)|\right] \\ &\le f_g(x,\bar{x},\theta)\|x-\bar{x}\|_1\end{aligned} \tag{5.29}$$

with $\theta = \theta_0$, where f_g is continuous and $\sup_{j,k\in\mathbb{N}} \mathbb{E}|f_g(I_j, I_k, \theta_0)|^{4+\varepsilon} + \mathbb{E}|f_g(I_1, \widetilde{I}_1, \theta_0)|^{4+\varepsilon} < \infty$. Moreover, there is a neighbourhood $U(\theta_0) \subseteq \Theta$ such that for all $\theta \in U(\theta_0)$, every element of the third derivative of g w.r.t. θ can be bounded by some function M with $\mathbb{E}M^2(I_1) < \infty$.

The parameter estimator is assumed to satisfy:

(M2) **(i)** The sequence of estimators $\widehat{\theta}_n$ admits the expansion

$$\widehat{\theta}_n - \theta_0 = \frac{1}{n}\sum_{k=1}^{n} l(X_k,\theta_0) + o_P(n^{-1/2}) \tag{5.30}$$

with $\mathbb{E}l(X_1,\theta_0) = 0$, $\mathbb{E}\|l(X_1,\theta_0)\|_1^{4+\varepsilon} < \infty$, and $X_k = (Y_k', I_k')'$, $k \in \mathbb{Z}$.

(ii) Moreover, $\|l(x,\theta_0) - l(\bar{x},\theta_0)\|_1 \le f_l(x,\bar{x},\theta_0)\,\|x-\bar{x}\|_1$, where the function $f_l(\cdot,\cdot,\theta_0)$ is symmetric, continuous and $\sup_{j,k\in\mathbb{N}} \mathbb{E}\max_{a\in[-A,A]^{d(p+1)+m\bar{p}}} |f_l(X_j + a, X_k, \theta_0)|^{4+\varepsilon} < \infty$ as well as $\mathbb{E}\max_{a\in[-A,A]^{d(p+1)+m\bar{p}}} |f_l(X_1+a, \widetilde{X}_1, \theta_0)|^{4+\varepsilon} < \infty$ for an independent copy $\widetilde{X}_1$ of X_1 and some $\varepsilon, A > 0$.

Finally, we make the following assumptions concerning the kernel function:

(M3) The entries of $K = diag(k_1,\ldots,k_d)$, $k_i : \mathbb{R}^{dp+m\bar{p}} \times \mathbb{R}^{dp+m\bar{p}} \times \Theta \to \mathbb{R}$, are symmetric in their first two arguments. The functions k_i are three times continuously differentiable w.r.t. θ and $k_i = k_i^{(0)}, \ldots, k_i^{(3)}$ are bounded and Lipschitz continuous in their first two arguments uniformly for all θ in some neighbourhood $U(\theta_0) \subseteq \Theta$.

We briefly comment on the assumptions. Condition (M1)(i) assures the existence of a unique stationary solution of (5.27) if additionally $\mathbb{E}\|\epsilon_1\|_1 + \mathbb{E}\|Z_1\|_1 < \infty$; see the proof of Lemma 5.11 for details. Moreover, it ensures the process $((Y_t', I_t')')_t$ to satisfy some weak dependence condition, namely the geometric-moment contraction condition GMC(1) of Shao and Wu [102]. For instance, Lipschitz contracting nonlinear AR(1) processes fulfil (M1)(i). Assumption (M1)(ii) is a joint constraint on smoothness and existence of moments of the derivatives of the conditional mean under the null hypothesis; a lack of smoothness has to be compensated by additional moments. In particular, functions g with continuous and bounded partial derivatives up to order three w.r.t. θ satisfy (M1) if their Lipschitz constant is sufficiently small. The condition (M2)(i) is a standard assumption regarding parameter estimators in the field of hypothesis testing, (M2)(ii) states some smoothness restrictions on the corresponding linearizing function. The set of feasible kernels defined in (M3) can be enlarged. However, since the kernel is not model inherent but chosen by hand, we restrict ourselves to this class for sake of technical simplification.

Before the asymptotic distributions of $\widehat{T}_n^{(u)}$ and $\widehat{T}_n^{(v)}$ are derived, we verify that these statistics can be approximated by degree-2 degenerate U- and V-statistics of $(X_t)_{t\in\mathbb{Z}}$ with $X_t = (Y_t', I_t')'$.

Lemma 5.11. *Suppose that $((Y_t', Z_t')')_{t\in\mathbb{Z}}$ is the stationary solution of equation (5.27) with $G(\cdot) = g(\cdot, \theta_0)$ satisfying $\mathbb{E}\|Z_1\|_1^2 + \mathbb{E}\|Y_1\|_1^{4+\varepsilon} < \infty$ for some $\varepsilon > 0$. Assume further that the assumptions (M1), (M2), and (M3) hold. Then, under $\mathcal{H}_0$,*

$$\widehat{T}_n^{(u)} = T_n^{(u)} + o_P(1) \quad \textit{and} \quad \widehat{T}_n^{(v)} = T_n^{(v)} + o_P(1)$$

with

$$T_n^{(u)} := \frac{1}{n} \sum_{\substack{j,k=1\\ j\neq k}}^{n} \epsilon_j' [K(I_j, I_k, \theta_0)\, \epsilon_k - 2\, V(I_j, \theta_0)\, l(X_k, \theta_0)] + [l(X_j, \theta_0)]'\, a(\theta_0)\, l(X_k, \theta_0),$$

$$T_n^{(v)} := \frac{1}{n} \sum_{j,k=1}^{n} \epsilon_j' [K(I_j, I_k, \theta_0)\, \epsilon_k - 2\, V(I_j, \theta_0)\, l(X_k, \theta_0)] + [l(X_j, \theta_0)]'\, a(\theta_0)\, l(X_k, \theta_0)$$

Here, $a(\theta_0)$ is a $(q \times q)$-matrix with entries

$$a(\theta_0)_{\alpha,\beta} = \mathbb{E}\left(\left[(g^{(1)}(I_1, \theta_0))' K(I_1, \widetilde{I}_1, \theta_0) g^{(1)}(\widetilde{I}_1, \theta_0)\right]_{\alpha,\beta}\right)$$

and $V(I_j, \theta_0)$ is a $(d \times q)$-matrix with entries

$$V(I_j, \theta_0)_{\alpha,\beta} = \mathbb{E}\left(k_\alpha(I_j, \widetilde{I}_j, \theta_0) g^{(1)}_{\alpha,\beta}(\widetilde{I}_j, \theta_0) | I_j\right),$$

where $\widetilde{I}_j$ denotes an independent copy of I_j, $j \in \mathbb{N}$.

Proof. First, note that $\mathbb{E}\|\varepsilon_1\|_1^{4+\varepsilon} = \mathbb{E}\|Y_1 - E(Y_1|I_1)\|_1^{4+\varepsilon} \leq C\, \mathbb{E}\|Y_1\|_1^{4+\varepsilon} < \infty$ under the assumptions of the lemma. According to Shao and Wu [102], Theorem 5.1, there exists a unique stationary solution $((Y_k', Z_k')')_{k\in\mathbb{Z}}$ to equation (5.27) with $G(\cdot) = g(\cdot, \theta_0)$ under $\mathcal{H}_0$ if the conditions of our Example 2.2.1 in Subsection 2.1.2 are satisfied. This in turn follows from assumption (M1)(i) and by defining the innovations as $\varepsilon_k := (\epsilon_k', Z_k')'$, $k \in \mathbb{Z}$. Moreover, their result implies that for all $n \in \mathbb{N}$ there is a random vector $(\widetilde{Y}_n', \widetilde{Z}_n')' \stackrel{d}{=} (Y_n', Z_n')'$ that is independent of $((Y_k', Z_k')')_{k \le 0}$ and that satisfies the GMC(1) condition, that is, $\mathbb{E}\|(\widetilde{Y}_n', \widetilde{Z}_n')' - (Y_n', Z_n')'\|_2 \le K\rho^n$ for some $K < \infty$ and $\rho \in (0,1)$. Moreover, the definition of the GMC(1) condition implies the existence of a copy $(\widetilde{Y}_n', \ldots, \widetilde{Y}_{n-P}', \widetilde{Z}_n', \ldots, \widetilde{Z}_{n-P}')'$ of the random vector $(Y_n', \ldots, Y_{n-P}', Z_n', \ldots, Z_{n-P}')'$ such that the first vector is independent of $((Y_k', Z_k')')_{k \le 0}$ for $n - P - 1 > 0$ and

$$\mathbb{E}\|(Y_n', \ldots, Y_{n-P}', Z_n', \ldots, Z_{n-P}')' - (\widetilde{Y}_n', \ldots, \widetilde{Y}_{n-P}', \widetilde{Z}_n', \ldots, \widetilde{Z}_{n-P}')'\|_2 \le K\,(P+1)\,\rho^{-P-1}\,\rho^n.$$

Due to the equivalence of norms on $\mathbb{R}^D$, the latter inequality implies that

$$\mathbb{E}\|X_n - \widetilde{X}_n\|_1 \le \sqrt{(d+m)(P+1)}\,K\,(P+1)\,\rho^{-P-1}\,\rho^n, \quad n \in \mathbb{N}, \tag{5.31}$$

where $X_n = (Y_n', I_n')'$ and $\widetilde{X}_n = (\widetilde{Y}_n', \widetilde{I}_n')'$ with $\widetilde{I}_n := (\widetilde{Y}_{n-p}', \ldots, \widetilde{Y}_{n-1}', Z_{n-\bar{p}+1}', \ldots, Z_n')'$. To sum up, $(X_k)_{k \in \mathbb{Z}}$ is τ-weakly dependent with exponentially decaying coefficients.

We restrict ourselves to the approximation of the statistic $\widehat{T}_n^{(v)}$ here. The calculations for $\widehat{T}_n^{(u)}$ can be carried out in complete analogy. Let

$$\begin{aligned}\widetilde{T}_n^{(v)} := \frac{1}{n}\sum_{j,k=1}^{n} \big\{&\epsilon_j'\,K(I_j, I_k, \theta_0)\,\epsilon_k - 2\,\epsilon_j'\,K(I_j, I_k, \theta_0)\,g^{(1)}(I_k, \theta_0)\,(\widehat{\theta}_n - \theta_0)\\ &+ (\widehat{\theta}_n - \theta_0)'\left[g^{(1)}(I_j, \theta_0)\right]'K(I_j, I_k, \theta_0)\,g^{(1)}(I_k, \theta_0)\,(\widehat{\theta}_n - \theta_0)\big\}.\end{aligned}$$

<u>Step 1:</u> $\widehat{T}_n^{(v)} = \widetilde{T}_n^{(v)} + o_P(1)$.
Elementwise Taylor expansion of $g(\cdot, \widehat{\theta}_n)$ and $K(\cdot, \cdot, \widehat{\theta}_n)$ results in

$$\widehat{T}_n^{(v)} - \widetilde{T}_n^{(v)} =$$

$$\frac{1}{n}\sum_{j,k=1}^{n}\sum_{i=1}^{d} \epsilon_{j,i}\,[(\widehat{\theta}_n - \theta_0)'\,k_i^{(1)}(I_j, I_k, \bar{\theta}_{n,i,j,k})]\,\epsilon_{k,i} \tag{5.32}$$

$$-\frac{1}{n}\sum_{j,k=1}^{n}\sum_{i=1}^{d} \epsilon_{j,i}\,k_i(I_j, I_k, \theta_0)\,[(\widehat{\theta}_n - \theta_0)'\,g_i^{(2)}(X_k, \bar{\theta}_{n,i,k})\,(\widehat{\theta}_n - \theta_0)] \tag{5.33}$$

$$\begin{aligned}-\frac{1}{n}\sum_{j,k=1}^{n}\sum_{i=1}^{d} &\epsilon_{j,i}\,[k_i^{(1)}(I_j, I_k, \bar{\theta}_{n,i,j,k})\,(\widehat{\theta}_n - \theta_0)]\\ &\times(\widehat{\theta}_n - \theta_0)'\left[2\,g_i^{(1)}(I_k, \theta_0) + g_i^{(2)}(I_k, \bar{\theta}_{n,i,k})\,(\widehat{\theta}_n - \theta_0)\right]\end{aligned} \tag{5.34}$$

$$\begin{aligned}+\frac{1}{n}\sum_{j,k=1}^{n}\sum_{i=1}^{d}&(\widehat{\theta}_n - \theta_0)'\left[g_i^{(1)}(I_j, \theta_0) + \frac{1}{2}g_i^{(2)}(I_j, \bar{\theta}_{n,i,j})\,(\widehat{\theta}_n - \theta_0)\right]\\ &\times k_i(I_j, I_k, \theta_0)\,[(\widehat{\theta}_n - \theta_0)'\,g_i^{(2)}(I_k, \bar{\theta}_{n,i,k})\,(\widehat{\theta}_n - \theta_0)]\end{aligned} \tag{5.35}$$

$$\begin{aligned}+\frac{1}{n}\sum_{j,k=1}^{n}\sum_{i=1}^{d}&(\widehat{\theta}_n - \theta_0)'\left[g_i^{(1)}(I_j, \theta_0) + \frac{1}{2}g_i^{(2)}(I_j, \bar{\theta}_{n,i,j})\,(\widehat{\theta}_n - \theta_0)\right]\\ &\times[(\widehat{\theta}_n - \theta_0)'\,k_i^{(1)}(I_j, I_k, \bar{\theta}_{n,i,j,k})]\,(\widehat{\theta}_n - \theta_0)'\left[g_i^{(1)}(I_k, \theta_0) + \frac{1}{2}g_i^{(2)}(I_k, \bar{\theta}_{n,i,k})\,(\widehat{\theta}_n - \theta_0)\right]\end{aligned} \tag{5.36}$$

for some random $\bar{\theta}_{n,i,j}$ and $\bar{\theta}_{n,i,j,k}$ between $\widehat{\theta}_n$ and θ_0. For sake of notational simplicity let $q = 1$ in the sequel. The extension of the approximations to higher dimensions is straightforward.

To verify asymptotic negligibility of the term (5.32), we expand $k_i^{(1)}(I_j, I_k, \bar{\theta}_{n,i,j,k}) = k_i^{(1)}(I_j, I_k, \theta_0) + k_i^{(2)}(I_j, I_k, \theta_0)(\bar{\theta}_{n,i,j,k} - \theta_0) + k_i^{(3)}(I_j, I_k, \bar{\bar{\theta}}_{n,i,j,k})(\bar{\theta}_{n,i,j,k} - \theta_0)^2/2$ for some random $\bar{\bar{\theta}}_{n,i,j,k}$ with $|\bar{\bar{\theta}}_{n,i,j,k} - \theta_0| \leq |\widehat{\theta}_n - \theta_0|$. Thus, the absolute value of (5.32) can be bounded from above by

$$o_P(1) \sum_{i=1}^{d} \left[\left| \frac{1}{n} \sum_{1 \leq j < k \leq n} \epsilon_{j,i}\, k_i^{(1)}(I_j, I_k, \theta_0)\, \epsilon_{k,i} \right| + \left| \frac{1}{n} \sum_{1 \leq j < k \leq n} \epsilon_{j,i}\, k_i^{(2)}(I_j, I_k, \theta_0)\, \epsilon_{k,i} \right| \right] + o_P(1)$$

as $\widehat{\theta}_n - \theta_0 = o_P(1)$ in view of Lemma 5.8. For every i the expressions within the absolute-value signs form degenerate U-statistics of the random variables $(X_k)_k$. Both U-statistics have finite second moments, which implies that they are of order $O_P(1)$, since the prerequisites of Lemma 3.7 are fulfilled due to assumption (M1)(i), (M3), and the moment constraints regarding Y_1. The required τ-dependence of the sample $(X_k)_k$ with $X_k = (Y_k', I_k')'$ follows from inequality (5.31).

In order to approximate the expression (5.33), we use a Taylor expansion of $g_i^{(2)}$ for $i = 1, \ldots, d$. This yields

$$\begin{aligned} |(5.33)| &= O_P(1) \left| \frac{1}{n^2} \sum_{j,k=1}^{n} \sum_{i=1}^{d} \epsilon_{j,i}\, k_i(I_j, I_k, \theta_0) \left[g_i^{(2)}(I_k, \theta_0) + g_i^{(3)}(I_k, \bar{\bar{\theta}}_{n,i,k})\, (\bar{\theta}_{n,i,k} - \theta_0) \right] \right| \\ &= o_P(1) + O_P(n^{-1/2}) \left| \frac{1}{n^{3/2}} \sum_{j,k=1}^{n} \sum_{i=1}^{d} \epsilon_{j,i}\, k_i(I_j, I_k, \theta_0)\, g_i^{(2)}(I_k, \theta_0) \right| \end{aligned}$$

for some $\bar{\bar{\theta}}_{n,i,k}$ between $\widehat{\theta}_n$ and θ_0. It is now sufficient to show that the remaining multiple sum is of order $O_P(1)$. For this purpose, we consider

$$\begin{aligned} &\frac{1}{n^3} \sum_{j,k,l,m=1}^{n} \mathbb{E} \left\{ \epsilon_{j,i} k_i(I_j, I_k, \theta_0) g_i^{(2)}(I_k, \theta_0) \right\} \left\{ \epsilon_{l,i} k_i(I_l, I_m, \theta_0) g_i^{(2)}(I_m, \theta_0) \right\} \\ &\leq O(1) + \frac{2}{n^3} \sum_{1 \leq j,k,m < l \leq n} \left| \mathbb{E} \left\{ \epsilon_{j,i} k_i(I_j, I_k, \theta_0) g_i^{(2)}(I_k, \theta_0) \right\} \left\{ \epsilon_{l,i} k_i(I_l, I_m, \theta_0) g_i^{(2)}(I_m, \theta_0) \right\} \right| \\ &\quad + \frac{2}{n^3} \sum_{1 \leq j < k,l,m \leq n} \left| \mathbb{E} \left\{ \epsilon_{j,i} k_i(I_j, I_k, \theta_0) g_i^{(2)}(I_k, \theta_0) \right\} \left\{ \epsilon_{l,i} k_i(I_l, I_m, \theta_0) g_i^{(2)}(I_m, \theta_0) \right\} \right| \\ &\quad + \frac{2}{n^3} \sum_{1 \leq k < j < l < m \leq n} \left| \mathbb{E} \left\{ \epsilon_{j,i} k_i(I_j, I_k, \theta_0) g_i^{(2)}(I_k, \theta_0) \right\} \left\{ \epsilon_{l,i} k_i(I_l, I_m, \theta_0) g_i^{(2)}(I_m, \theta_0) \right\} \right| \\ &\quad + \frac{2}{n^3} \sum_{1 \leq m < j < l < k \leq n} \left| \mathbb{E} \left\{ \epsilon_{j,i} k_i(I_j, I_k, \theta_0) g_i^{(2)}(I_k, \theta_0) \right\} \left\{ \epsilon_{l,i} k_i(I_l, I_m, \theta_0) g_i^{(2)}(I_m, \theta_0) \right\} \right|. \end{aligned}$$

It can easily be seen that the first two sums are uniformly bounded in n. To this end, we use a coupling with appropriate copies $\widetilde{X}_l$ and $(\widetilde{X}_k', \widetilde{X}_l', \widetilde{X}_m')'$ of X_l and $(X_k', X_l', X_m')'$ that are independent of $(X_j', X_k', X_m')'$ and X_j, respectively, if $l-(P+1)-\max\{j,k,m\}>0$ and $\min\{k,l,m\}-(P+1)-j>0$, respectively. Note that in these cases

$$\mathbb{E}\left\{\epsilon_{j,i}\, k_i(I_j, I_k, \theta_0)\, g_i^{(2)}(I_k,\theta_0)\right\}\left\{[\widetilde{Y}_{l,i}-g_i(\widetilde{I}_l,\theta_0)]\, k_i(\widetilde{I}_l, I_m,\theta_0)\, g_i^{(2)}(I_m,\theta_0)\right\}=0$$

and

$$\mathbb{E}\left\{\epsilon_{j,i}\, k_i(I_j, \widetilde{I}_k, \theta_0)\, g_i^{(2)}(\widetilde{I}_k,\theta_0)\right\}\left\{[\widetilde{Y}_{l,i}-g_i(\widetilde{I}_l,\theta_0)]\, k_i(\widetilde{I}_l, \widetilde{I}_m,\theta_0)\, g_i^{(2)}(\widetilde{I}_m,\theta_0)\right\}=0.$$

The desired order of the corresponding summands can then be obtained under (M1) and (M3) invoking the usual arguments.

In order to show uniform boundedness of the third sum, we introduce a vector $(\widetilde{X}_l', \widetilde{X}_m')' \stackrel{d}{=} (X_l', X_m')'$ that is independent of $(X_k', X_j')'$ and that satisfies the condition (5.31) with $n=l-j$ as long as $l-(P+1)-j>0$. This leads to

$$\begin{aligned}
&\frac{2}{n^3}\sum_{1\le k<j<l<m\le n}\left|\mathbb{E}\left\{\epsilon_{j,i}\, k_i(I_j, I_k,\theta_0)\, g_i^{(2)}(I_k,\theta_0)\right\}\left\{\epsilon_{l,i}\, k_i(I_l, I_m,\theta_0)\, g_i^{(2)}(I_m,\theta_0)\right\}\right| \\
&\le O(1)+\frac{1}{n}\left|\frac{1}{n}\sum_{\substack{j,k=1\\ j\neq k}}^{n}\mathbb{E}\epsilon_{j,i}\, k_i(I_j, I_k,\theta_0)\, g_i^{(2)}(I_k,\theta_0)\right|^2 .
\end{aligned}$$

The expression within the absolute value signs is of order $O(1)$ as, in case of $j>k+P+1$,

$$\begin{aligned}
\left|\mathbb{E}\epsilon_{j,i}\, k_i(I_j, I_k,\theta_0)\, g_i^{(2)}(I_k,\theta_0)\right| &\le C\,\mathbb{E}|\epsilon_{j,i}\, k_i(I_j, I_k,\theta_0)-\widetilde{\epsilon}_{j,i}\, k_i(\widetilde{I}_j, I_k,\theta_0)|\,|g_i^{(2)}(I_k,\theta_0)| \\
&\le C\,\rho^{(j-k)\delta}
\end{aligned}$$

for a $\delta>0$, some appropriate $\widetilde{X}_j \stackrel{d}{=} X_j$ that is independent of X_k, and $\widetilde{\epsilon}_{j,i} := \widetilde{Y}_j - g(\widetilde{I}_j,\theta_0)$. Similarly, one gets $|\mathbb{E}\epsilon_{j,i}\, k_i(I_j, I_k,\theta_0)\, g_i^{(2)}(I_k,\theta_0)| \le C\,\rho^{(k-j)\delta}$ when $k>j+P+1$ by introducing an suitable copy $\widetilde{I}_k$ of I_k that is independent of X_j. Eventually, we are to verify

$$\frac{2}{n^3}\sum_{1\le m<j<l<k\le n}\left|\mathbb{E}\left\{\epsilon_{j,i}\, k_i(I_j, I_k,\theta_0)\, g_i^{(2)}(I_k,\theta_0)\right\}\left\{\epsilon_{l,i}\, k_i(I_l, I_m,\theta_0)\, g_i^{(2)}(I_m,\theta_0)\right\}\right| = O(1).$$

Under (M1) and (M3) it suffices to consider those summands such that the minimal gap between the corresponding indices is larger than $P+1$. Denote this set of indices by J. Invoking the dependence structure of the underlying sample and the continuity properties of the involved functions in the usual manner, the introduction of copies $\widetilde{I}_k$ of I_k that are independent of $(X_m', X_j', X_l')'$ imply that the above sum can be bounded from above by

$$\frac{2}{n^3}\sum_{\substack{1\le m<j<l<k\le n\\ (m,j,l,k)\in J}}\left|\mathbb{E}\left\{\epsilon_{j,i}\, k_i(I_j, \widetilde{I}_k,\theta_0)\, g_i^{(2)}(\widetilde{I}_k,\theta_0)\right\}\left\{\epsilon_{l,i}\, k_i(I_l, I_m,\theta_0)\, g_i^{(2)}(I_m,\theta_0)\right\}\right| + O(1).$$

Proceeding iteratively in complete analogy with the remaining indices l, j and m, i.e. introducing a coupling of the variable with the largest index, we end up with the upper estimate

$$O(1) + \frac{2}{n} \sum_{\substack{1 \leq m < j \leq n \\ (j,m) \in J}} \left| \mathbb{E} \int_{\mathbb{R}^{dp+m\bar{p}}} \epsilon_{j,i}\, k_i(I_j, y, \theta_0)\, g_i^{(2)}(y, \theta_0) P_{I_1}(dy) \right|$$

$$\times \left| \mathbb{E} \int_{\mathbb{R}^{d(p+1)+m\bar{p}}} [z_1 - g_i(z_2, \theta_0)]\, k_i(z_2, I_m, \theta_0)\, g_i^{(2)}(I_m, \theta_0) P_{X_1}(dz) \right|,$$

where $z = (z_1', z_2')'$ with $z_1 \in \mathbb{R}^d$ and $z_2 \in \mathbb{R}^{dp+m\bar{p}}$. Note that all summands of the remaining sum are equal to zero. Consequently, we finally achieve the desired order of (5.33).

Similar arguments as in the considerations of (5.32) and (5.33) lead to (5.34) + (5.35) + (5.36) $= o_P(1)$ due to the moment constraints concerning $g_i^{(1)}, g_i^{(2)}$, and $g_i^{(3)}, i = 1, \ldots, d$. Thus, $\widehat{T}_n^{(v)} - \widetilde{T}_n^{(v)} = o_P(1)$.

<u>Step 2:</u> $\widetilde{T}_n^{(v)} = T_n^{(v)} + o_P(1)$.

For sake of notational simplicity all calculations are only stated for $q = 1$ again. According to the representation (5.30) of $\widehat{\theta}_n - \theta$, we obtain

$$\begin{aligned}
&\frac{1}{n} \sum_{j,k=1}^{n} \epsilon_j'\, K(I_j, I_k, \theta_0)\, g^{(1)}(I_k, \theta_0)\, (\widehat{\theta}_n - \theta_0) \\
&= \frac{1}{n^2} \sum_{j,k,m=1}^{n} \sum_{i=1}^{d} \epsilon_{j,i}\, k_i(I_j, I_k, \theta_0)\, g_i^{(1)}(I_k, \theta_0)\, l(X_m, \theta_0) \\
&\quad + o_P(n^{-1/2})\, \frac{1}{n} \sum_{j,k=1}^{n} \sum_{i=1}^{d} \epsilon_{j,i}\, k_i(I_j, I_k, \theta_0)\, g_i^{(1)}(I_k, \theta_0),
\end{aligned}$$

where the latter summand vanishes asymptotically, cf. the analysis of the quantity (5.33). Moreover,

$$\begin{aligned}
&\frac{1}{n^2} \sum_{j,k,m=1}^{n} \epsilon_j' \left[K(I_j, I_k, \theta_0)\, g^{(1)}(I_k, \theta_0) - V(I_j, \theta_0) \right] l(X_m, \theta_0) \\
&= O_P(1)\, \frac{1}{n^{3/2}} \sum_{j,k=1}^{n} \left\{ \epsilon_j' \left[K(I_j, I_k, \theta_0)\, g^{(1)}(I_k, \theta_0) - V(I_j, \theta_0) \right] \right. \\
&\qquad\qquad \left. + \epsilon_k' \left[K(I_k, I_j, \theta_0)\, g^{(1)}(I_j, \theta_0) - V(I_k, \theta_0) \right] \right\}.
\end{aligned}$$

The double sum builds a degenerate V-statistic multiplied with $\sqrt{n}$. According to the continuity assumptions on the involved functions and Lemma 3.7, this quantity vanishes asymptotically. Summing up, we obtain

$$\frac{1}{n} \sum_{j,k=1}^{n} \epsilon_j'\, K(I_j, I_k, \theta_0)\, g^{(1)}(I_k, \theta_0)\, (\widehat{\theta}_n - \theta_0) = \frac{1}{n} \sum_{j,k=1}^{n} \epsilon_j'\, V(I_j, \theta_0)\, l(X_k, \theta_0) + o_P(1).$$

It remains to consider

$$\begin{aligned}
&\frac{1}{n}\sum_{j,k=1}^{n}(\widehat{\theta}_n-\theta_0)\,\big[g^{(1)}(I_j,\theta_0)\big]'\,K(I_j,I_k,\theta_0)\,g^{(1)}(I_k,\theta_0)\,(\widehat{\theta}_n-\theta_0)\\
&=\frac{1}{n^3}\sum_{i,j,k,m=1}^{n} l(X_i,\theta_0)\,\big[g^{(1)}(I_j,\theta_0)\big]'\,K(I_j,I_k,\theta_0)\,g^{(1)}(I_k,\theta_0)\,l(X_m,\theta_0)\\
&\quad+o_P(1)\left[\frac{1}{n^2}\sum_{j,k=1}^{n}\big[g^{(1)}(I_j,\theta_0)\big]'\,K(I_j,I_k,\theta_0)\,g^{(1)}(I_k,\theta_0)\right]\left[\frac{1}{\sqrt{n}}\sum_{m=1}^{n} l(X_m,\theta_0)\right]+o_P(1),\\
&=\frac{1}{n^3}\sum_{i,j,k,m=1}^{n} l(X_i,\theta_0)\,\big[g^{(1)}(I_j,\theta_0)\big]'\,K(I_j,I_k,\theta_0)\,g^{(1)}(I_k,\theta_0)\,l(X_m,\theta_0)+o_P(1).
\end{aligned}$$

The latter equality is obtained by applying Lemma 5.8 to $n^{-1/2}\sum_{m=1}^{n} l(X_m,\theta_0)$ and by virtue of the assumptions (M1) to (M3). Thus, we still have to prove that $T_n := n^{-2}\sum_{j,k=1}^{n}[g^{(1)}(I_j,\theta_0)]'\,K(I_j,I_k,\theta_0)\,g^{(1)}(I_k,\theta_0)-a(\theta_0)\xrightarrow{P}0$. To this end, a Hoeffding decomposition of the kernel associated with the above V-statistic is invoked as follows:

$$T_n=\frac{2}{n}\sum_{j=1}^{n} h_1(I_j)+\frac{1}{n^2}\sum_{j,k=1}^{n} h_2(I_j,I_k),$$

where $h_1(x) := \mathbb{E}[g^{(1)}(x,\theta_0)]'K(x,I_1,\theta_0)g^{(1)}(I_1,\theta_0)+a(\theta_0)$ and

$$h_2(x,y)=[g^{(1)}(x,\theta_0)]'K(x,y,\theta_0)g^{(1)}(y,\theta_0)+a(\theta_0)-h_1(x)-h_1(y),\quad x,y\in\mathbb{R}^{dp+m\bar{p}}.$$

Note that h_2 is degenerate. Due to Lemma 3.7 the corresponding double sum tends to zero in probability. Moreover, the introduction of random variables $\widetilde{I}_k \overset{d}{=} I_k$, that are chosen independently of I_j and such that (5.31) holds with $n=k-j$, yields

$$\frac{1}{n^2}\sum_{j,k=1}^{n}\mathbb{E}h_1(I_j)\,h_1(I_k)\le o(1)+\frac{2}{n^2}\sum_{1\le j<j+P+1<k\le n}\mathbb{E}h_1(I_j)\,[h_1(I_k)-h_1(\widetilde{I}_k)]=o(1),$$

which eventually completes the proof of the lemma. □

This assertion enables us to apply Theorem 3.2 in order to deduce the limiting distributions of $T_n^{(u)}$ and $T_n^{(v)}$, and thus of $\widehat{T}_n^{(u)}$ and $\widehat{T}_n^{(v)}$, respectively.

Theorem 5.3. *Suppose that the prerequisites of Lemma 5.11 are satisfied. Then,*

$$\begin{aligned}
&\widehat{T}_n^{(u)}\xrightarrow{d} Z_M\quad\textit{and}\\
&\widehat{T}_n^{(v)}\xrightarrow{d} Z_M+\mathbb{E}\left\{\epsilon_1'\,[K(I_1,I_1,\theta_0)\,\epsilon_1-2\,V(I_1,\theta_0)\,l(X_1,\theta_0)]+[l(X_1,\theta_0)]'\,a(\theta_0)\,l(X_1,\theta_0)\right\}
\end{aligned}$$

under $\mathcal{H}_0$. Here, the distribution of Z_M is the asymptotic distribution of $n\,U_n$ defined in Theorem 3.2, where U_n denotes a U-statistic based on the underlying sample and whose kernel function is given by

$$\begin{aligned}
h(x,y,\theta_0) := {}&[x_1-g(x_2,\theta_0)]'\,K(x_2,y_2,\theta_0)\,[y_1-g(y_2,\theta_0)]+[l(x,\theta_0)]'\,a(\theta_0)\,l(y,\theta_0)\\
&-[V(y_1,\theta_0)\,l(x,\theta_0)]'\,[y_1-g(y_2,\theta_0)]-[x_1-g(x_2,\theta_0)]'\,V(x_2,\theta_0)\,l(y,\theta_0)
\end{aligned}$$

with $x=(x_1',x_2')'$, $y=(y_1',y_2')'$, $x_1,y_1\in\mathbb{R}^d$ and $x_2,y_2\in\mathbb{R}^{dp+m\bar{p}}$.

Proof. As verified already at the beginning of the previous proof, the underlying sample satisfies the condition (A1). According to Lemma 5.11 it suffices to consider the expressions $T_n^{(u)}$ and $T_n^{(v)}$. Under the null hypothesis, these are degenerate U- and V-statistics multiplied with the sample size and whose kernel function is equivalent to the one stated in the theorem. The function h satisfies the condition (A2) which can easily be verified by applying Hölder's inequality. Since we shall invoke Theorem 3.2 in order to determine the asymptotics of $T_n^{(u)}$ and $T_n^{(v)}$, it remains to check assumption (A4). In view of (M1) to (M3), the following relation holds:

$$
\begin{aligned}
&|h(x,y,\theta_0) - h(\bar{x},\bar{y},\theta_0)| \\
&\leq C\Big\{1 + \|x_1\|_1 + \|\bar{x}_1\|_1 + \|g(x_2,\theta_0)\|_1 + \|g(\bar{x}_2,\theta_0)\|_1 + H(x_2,\bar{x}_2,\theta_0) + H(\bar{x}_2,x_2,\theta_0) \\
&\quad + \|l(x,\theta_0)\|_1 + \|l(\bar{x},\theta_0)\|_1 + f_l(x,\bar{x},\theta_0)\Big\}\Big\{1 + \|y_1\|_1 + \|\bar{y}_1\|_1 + \|g(y_2,\theta_0)\|_1 \\
&\quad + \|g(\bar{y}_2,\theta_0)\|_1 + H(y_2,\bar{y}_2,\theta_0) + H(\bar{y}_2,y_2,\theta_0) + \|l(y,\theta_0)\|_1 + \|l(\bar{y},\theta_0)\|_1 + f_l(y,\bar{y},\theta_0)\Big\} \\
&\quad \times (\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1).
\end{aligned}
$$

Thus, there is a continuous and symmetric function f_{θ_0} such that

$$
|h(x,y,\theta_0) - h(\bar{x},\bar{y},\theta_0)| \leq f_{\theta_0}(x,\bar{x},y,\bar{y})\,[\|x-\bar{x}\|_1 + \|y-\bar{y}\|_1]
$$

with

$$
\sup_{k_1,\dots,k_5\in\mathbb{N}} \mathbb{E}\left\{\max_{a_1,a_2\in[-A,A]^{d(p+1)+m\bar{p}}} [f_{\theta_0}(\bar{X}_{k_1},\bar{X}_{k_2}+a_1,\bar{X}_{k_3},\bar{X}_{k_4}+a_2)]^{1+\mu}\|\bar{X}_{k_1}\|_1\right\} < \infty
$$

for some $\mu > 0$ and any $(\bar{X}_{k_1}',\dots,\bar{X}_{k_5}')'$ consisting of independent subvectors $(\bar{X}_{k_{j_1(m)}}',\dots,\bar{X}_{k_{j_l(m)}}')' \stackrel{d}{=} (X_{k_{j_1(m)}}',\dots,X_{k_{j_l(m)}}')'$, $l,m = 1,\dots,5$. Consequently, the conditions (A1), (A2), as well as (A4) are satisfied and Theorem 3.2 applies. □

In what follows, we study the behaviour of the test statistics under the alternative hypothesis $\mathcal{H}_1$. It turns out after some approximation steps that in order to verify asymptotic unboundedness of the statistics $\widehat{T}_n^{(u)}$ and $\widehat{T}_n^{(v)}$, the constraint

$$
\Delta(\theta_0) = \mathbb{E}\left([\mathbb{E}(Y_1|I_1) - g(I_1,\theta_0)]'K(I_1,\widetilde{I}_1,\theta_0)[\mathbb{E}(\widetilde{Y}_1|\widetilde{I}_1) - g(\widetilde{I}_1,\theta_0)]\right) > 0
$$

has to be satisfied. Here, $(\widetilde{Y}_1',\widetilde{I}_1')'$ denotes an independent copy of $(Y_1',I_1')'$. Three different sufficient conditions concerning the kernel matrix are stated below. All of these constraints are based on the so-called integrated conditional moment approach, that is, they assure equivalence between vanishing conditional moments, $\mathbb{E}(Y_1 - g(I_1,\theta_0)|I_1) = 0$ a.s., and unconditional moments, $\mathbb{E}[(Y_1 - g(I_1,\theta_0))'f(I_1,x)] = 0$ for a certain function f and almost all x in a compact subset of $\mathbb{R}^v$. However, this idea does not become apparent from the conditions themselves, which are far from intuitive, but from the proof of the subsequent result.

Lemma 5.12. *Suppose that $(Y_t)_t$ is stationary solution of (5.27), where G satisfies the conditions (I) and (II) of Subsection 4.4.4 with $\varepsilon_k = (\epsilon_t', Z_t')'$. Moreover, let $\mathbb{E}\|Y_1\|_1^{4+\varepsilon} + \mathbb{E}\|Z_1\|_1^2 < \infty$ and $\sqrt{n}(\widehat{\theta}_n - \theta_0) = O_P(1)$ for some $\theta_0 \in \Theta$ with $\mathbb{E}\|g(I_1, \theta_0)\|_1^{4+\varepsilon} < \infty$. Furthermore, assume that (M1), (M3), and one of the following conditions hold true with some bounded Lipschitz continuous one-to-one mappings $\zeta_i : \mathbb{R}^{dp+m\bar{p}} \to \mathbb{R}^{dp+m\bar{p}}$, $i = 1, \ldots, d$:*

(i) The diagonal terms of $K(\cdot, \cdot, \theta_0)$ are absolutely Lebesgue integrable and admit $k_i(x, y, \theta_0) = \bar{k}_{i,\theta_0}(\zeta_i(x) - \zeta_i(y))$, $i = 1, \ldots, d$. The Fourier transforms $\mathcal{F}\bar{k}_{i,\theta_0}$ of $\bar{k}_{i,\theta_0}$ are nonnegative and do not vanish in a neighbourhood of the origin.

(ii) The diagonal terms of $K(\cdot, \cdot, \theta_0)$ are absolutely Lebesgue integrable and admit $k_i(x, y, \theta_0) = \bar{k}_{i,\theta_0}(x - y)$, $i = 1, \ldots, d$. The Fourier transforms $\mathcal{F}\bar{k}_{i,\theta_0}$ of $\bar{k}_{i,\theta_0}$ are positive a.e. w.r.t. the Lebesgue measure.

(iii) The diagonal elements of $K(\cdot, \cdot, \theta_0)$ have the following representation

$$k_i(x, y, \theta_0) = \int_{\mathbb{R}^{dp+m\bar{p}+1}} W_i((1, [\zeta_i(x)]')'t)\, W_i((1, [\zeta_i(y)]')'t)\, w_i^2(t, \theta_0)\, dt,$$

where $W_i : \mathbb{R} \to \mathbb{R}$ are analytic, non-polynomial functions. The weight functions $w_i : \mathbb{R}^{dp+m\bar{p}+1} \times \Theta \to \mathbb{R}$ are assumed to be measurable and to satisfy

$$0 < \int_{\mathbb{R}^{dp+m\bar{p}+1}} \sup_{x,y} |W_i((1, [\zeta_i(x)]')'t)\, W_i((1, [\zeta_i(y)]')'t)| w_i^2(t, \theta_0)\, dt < \infty.$$

Then, under any fixed alternative in $\mathcal{H}_1$,

$$P\left(\widehat{T}_n^{(u)} > K\right) \underset{n\to\infty}{\longrightarrow} 1 \quad \text{and} \quad P\left(\widehat{T}_n^{(v)} > K\right) \underset{n\to\infty}{\longrightarrow} 1, \quad \forall\, K \in \mathbb{R}.$$

Since the conditions are rather technical, we give some examples before proving the results. Let $\zeta(x) = (\arctan(x_1), \ldots, \arctan(x_{dp+m\bar{p}}))'$. Assumptions (i) and (ii) of the foregoing lemma are satisfied for instance by the following frequently used kernels:

- Gauss kernel $\bar{k}_i(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ with $\mathcal{F}(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$,
- Cauchy kernel $\bar{k}_i(x) = \frac{1}{\pi(1+x^2)}$ with $\mathcal{F}(t) = \frac{1}{\sqrt{2\pi}} e^{-|t|}$,
- Triangular kernel $\bar{k}_i(x) = (1 - |x|)\mathbb{1}_{[-1,1]}(x)$ with $\mathcal{F}(t) = \frac{2(1-\cos t)}{\sqrt{2\pi} t^2}$,
- Picard kernel $\bar{k}_i(x) = \frac{1}{2} e^{-|x|}$ with $\mathcal{F}(t) = \frac{1}{\sqrt{2\pi}(1+t^2)}$.

They are violated when the uniform or the Epanechnikov kernel are applied.

The third constraint holds true for square integrable functions w with bounded support if e.g. $W(y) = e^y$, which was applied by Bierens [14], or the logistic function $W(y) = 1/[1 + e^{c-y}]$, $c \neq 0$, are used.

Proof. Step 1: $n^{-1}\,\widehat{T}_n^{(v)} \xrightarrow{P} \mathbb{E}([\mathbb{E}(Y_1|I_1) - g(I_1,\theta_0)]' K(I_1,\widetilde{I}_1,\theta_0)[\mathbb{E}(Y_1|\widetilde{I}_1) - g(\widetilde{I}_1,\theta_0)])$.
According to the assumptions on the set of functions g under $\mathcal{H}_0$, on the sequence of parameter estimators $(\widehat{\theta}_n)_n$ and on the kernel function k we obtain

$$
\begin{aligned}
n^{-1}\,\widehat{T}_n^{(v)} &= \frac{1}{n^2}\sum_{j,k=1}^{n}[Y_j - \mathbb{E}(Y_j|I_j)]' K(I_j,I_k,\theta_0)[Y_k - \mathbb{E}(Y_k|I_k)] \\
&\quad + \frac{2}{n^2}\sum_{j,k=1}^{n}[Y_j - \mathbb{E}(Y_j|I_j)]' K(I_j,I_k,\theta_0)[\mathbb{E}(Y_k|I_k) - g(I_k,\theta_0)] \\
&\quad + \frac{1}{n^2}\sum_{j,k=1}^{n}[\mathbb{E}(Y_j|I_j) - g(I_j,\theta_0)]' K(I_j,I_k,\theta_0)[\mathbb{E}(Y_k|I_k) - g(I_k,\theta_0)] + o_P(1).
\end{aligned}
$$

In view of Theorem 5.1 of Shao and Wu [102] the process $(X_k)_k$ has a Bernoulli shift representation with innovations $((\epsilon_k', Z_k')')_k$ and is τ-dependent with exponentially decaying coefficients. Since the first sum is a degenerate V-statistic in these variables, we obtain its asymptotic negligibility in analogy to the proof of Lemma 3.7. Even though the function G, determining $\mathbb{E}(Y_k|I_k) = G(I_k)$, does not satisfy the continuity assumptions of Lemma 3.7, its proof remains valid since $\mathbb{E}\|G(I_k) - G(\widetilde{I}_k)\|_1 \le C\rho^k$, for some $\rho \in (0,1)$ and a copy $\widetilde{I}_k$ of I_k that is independent of $(I_l)_{l\le 0}$ for $k > P+1$, can be employed instead. The latter inequality results from condition (II) and the fact that the vector $\widetilde{I}_k$ can be chosen such that its last component, $\widetilde{Z}_k$, coincides with the last component, Z_k, of I_k. Coupling methods, similar to those used to investigate the quantity (5.33) in the proof of Lemma 5.11, yield that the middle term is asymptotically negligible. Finally, a Hoeffding decomposition of the kernel of the remaining summand leads to

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{j,k=1}^{n}[\mathbb{E}(Y_j|I_j) - g(I_j,\theta_0)]' K(I_j,I_k,\theta_0)[\mathbb{E}(Y_k|I_k) - g(I_k,\theta_0)] \\
&= \Delta(\theta_0) + \frac{2}{n}\sum_{j=1}^{n} h_1(I_j) + \frac{1}{n^2}\sum_{j,k=1}^{n} h_2(I_j,I_k)
\end{aligned}
\tag{5.37}
$$

with

$$
\begin{aligned}
\Delta(\theta_0) &= \mathbb{E}([\mathbb{E}(Y_1|I_1) - g(I_1,\theta_0)]' K(I_1,\widetilde{I}_1,\theta_0)[\mathbb{E}(\widetilde{Y}_1|\widetilde{I}_1) - g(\widetilde{I}_1,\theta_0)]), \\
h_1(I_j) &= \mathbb{E}([\mathbb{E}(Y_j|I_j) - g(I_j,\theta_0)]'\, K(I_j,\widetilde{I}_j,\theta_0)\, [\mathbb{E}(\widetilde{Y}_j|\widetilde{I}_j) - g(\widetilde{I}_j,\theta_0)] \mid I_j) - \Delta(\theta_0), \\
h_2(I_j,I_k) &= [\mathbb{E}(Y_j|I_j) - g(I_j,\theta_0)]'\, K(I_j,I_k,\theta_0)\, [\mathbb{E}(Y_k|I_k) - g(I_k,\theta_0)] \\
&\quad - h_1(I_k) - h_1(I_j) - \Delta(\theta_0),
\end{aligned}
$$

where $(\widetilde{Y}_j', \widetilde{I}_j')'$ is an independent copy of $(Y_j', I_j')'$. Analogous arguments as at the end of proof of Lemma 5.11 imply that the middle summand of the r.h.s. of equation (5.37) is of order $o_P(1)$. The same order can be derived for the last sum which is a degenerate U-statistic in $(I_k)_k$. This finally completes the first step. Obviously, we also have $n^{-1}\,\widehat{T}_n^{(u)} \xrightarrow{P} \Delta(\theta_0)$.
Step 2: $\Delta(\theta_0) > 0$ *if (i) holds true.*

By inverse Fourier transform we obtain $\bar{k}_{j,\theta_0}(x) = (2\pi)^{-(dp+m\bar{p})/2} \int_{\mathbb{R}^{dp+m\bar{p}}} \mathcal{F}\bar{k}_{j,\theta_0}(t) e^{it'x} dt$, $j = 1, \ldots, d$. Thus, the application of Fubini's theorem leads to

$$\begin{aligned}\Delta(\theta_0) &= (2\pi)^{-(dp+m\bar{p})/2} \sum_{j=1}^{d} \int_{\mathbb{R}^{dp+m\bar{p}}} \mathbb{E}\left([\mathbb{E}(Y_1|I_1) - g(I_1,\theta_0)]_j \, e^{i\zeta_j(I_1)'t} \right) \\ &\qquad\qquad \times \mathbb{E}\left([\mathbb{E}(Y_1|I_1) - g(I_1,\theta_0)]_j \, e^{-i\zeta_j(I_1)'t} \right) \mathcal{F}\bar{k}_{i,\theta_0}(t)\, dt \\ &= (2\pi)^{-(dp+m\bar{p})/2} \sum_{j=1}^{d} \int_{\mathbb{R}^{dp+m\bar{p}}} \left| \mathbb{E}\left([\mathbb{E}(Y_1|I_1) - g(I_1,\theta_0)]_j \, e^{i\zeta_j(I_1)'t} \right) \right|^2 \mathcal{F}\bar{k}_{i,\theta_0}(t)\, dt,\end{aligned}$$

where the subscript j denotes the j^{th} element of the corresponding vector. By Theorem 1(II) of Bierens [13], the expression $\Delta(\theta_0)$ is strictly positive since $\mathbb{E}(Y_1 - g(I_1,\theta_0)|I_1) = \mathbb{E}(Y_1 - g(I_1,\theta_0)|\zeta_j(I_1))$ almost surely.

Step 3: $\Delta(\theta_0) > 0$ *if (ii) holds true.*

One proceeds as in the previous step but employs Theorem 1(I) of Bierens [13] instead of Theorem 1(II).

Step 4: $\Delta(\theta_0) > 0$ *if (iii) holds true.*

Fubini's theorem yields

$$\Delta(\theta_0) = \sum_{j=1}^{d} \int_{\mathbb{R}^{dp+m\bar{p}+1}} \left| \mathbb{E}\{ [\mathbb{E}(Y_1|I_1) - g(I_1,\theta_0)]_j \, W_j((1,[\zeta_j(I_1)]')'t) \, w_j(t,\theta_0)\} \right|^2 dt.$$

Thus, it remains to verify that there exists a $j \in \{1,\ldots,d\}$ such that $\mathbb{E}\{[\mathbb{E}(Y_1|I_1) - g(I_1,\theta_0)]_j \, W_j((1,[\zeta_j(I_1)]')'t) \, w_j(t,\theta_0)\} \neq 0$ on a set of positive Lebesgue measure, which follows from Theorem 2.3 of Stinchcombe and White [104]. □

Eventually, we consider the behaviour of both test statistics under local alternatives similar to those that were investigated by Escanciano [60] or by Bierens and Ploberger [15]:

$$\mathcal{H}_{1,n}: \quad Y_{n,t} = g(I_t,\theta_0) + \frac{A(I_t)}{\sqrt{n}} + \epsilon_t, \quad \text{a.s.} \quad \text{with} \quad P(A(I_t) \neq 0) > 0.$$

The following additional assumptions concerning the smoothness of the function A and the behaviour of the parameter estimator under $\mathcal{H}_{1,n}$ are imposed:

(M4) **(i)** The function $A : \mathbb{R}^{dp+m\bar{p}} \to \mathbb{R}^d$ is Lipschitz continuous and satisfies $\mathbb{E}\|A(I_t)\|_1^{4+\varepsilon} < \infty$ for some $\varepsilon > 0$.

(ii) Under $\mathcal{H}_{1,n}$, the parameter estimator $\widehat{\theta}_{n,n}$ admits the expansion

$$\widehat{\theta}_{n,n} - \theta_0 = \frac{T_a}{\sqrt{n}} + \frac{1}{n}\sum_{k=1}^{n} l(X_k,\theta_0) + o_P(n^{-1/2}),$$

for some $T_a \in \mathbb{R}^q$ and $X_k = (Y_k', I_k')'$.

(ii) The diagonal elements k_i of the kernel matrix K are of the form $k_i(x,y) = \int_{\mathbb{R}^\nu} \bar{k}_i(x,t)\, \bar{k}_i(y,t)\, w_i(t)\, dt$ for Lipschitz continuous functions $\bar{k}_i : \mathbb{R}^{dp+m\bar{p}} \times \mathbb{R}^\nu \to \mathbb{R}_+$ and a.s. nonnegative integrable functions $w_i : \mathbb{R}^\nu \to \mathbb{R}_+$.

For sake of simplicity we restrict ourselves to Lipschitz continuous alternatives A here. It easy to convince oneself that it suffices to postulate a weaker smoothness constraint, similar to the one in (M2) concerning the function l, instead. Analogously, the smoothness assumption regarding the diagonal elements of the kernel matrix K can be relaxed.

The corresponding test statistics are denoted by

$$\widehat{T}^{(u)}_{n,n} := \frac{1}{n} \sum_{\substack{j,k=1 \\ j\neq k}}^{n} \left[Y_{n,j} - g(I_j, \widehat{\theta}_{n,n})\right]' K(I_j, I_k, \widehat{\theta}_{n,n}) \left[Y_{n,k} - g(I_k, \widehat{\theta}_{n,n})\right]$$

and

$$\widehat{T}^{(v)}_{n,n} := \frac{1}{n} \sum_{j,k=1}^{n} \left[Y_{n,j} - g(I_j, \widehat{\theta}_{n,n})\right]' K(I_j, I_k, \widehat{\theta}_{n,n}) \left[Y_{n,k} - g(I_k, \widehat{\theta}_{n,n})\right].$$

Proposition 5.7. *Suppose that the assumptions (M1) to (M4) are satisfied and that the row-wise stationary triangular scheme of response variables $(Y_{n,k})_{k=1}^n$ under $\mathcal{H}_{1,n}$, $n \in \mathbb{N}$, satisfies $\sup_{n\in\mathbb{N}} \mathbb{E}\|Y_{n,1}\|_1^{4+\varepsilon} + \mathbb{E}\|Z_1\|_1^2 < \infty$. Then, as $n \to \infty$,*

$$\widehat{T}^{(u)}_{n,n} \xrightarrow{d} Z_{M,loc} \quad \text{and} \quad \widehat{T}^{(v)}_{n,n} \xrightarrow{d} Z_{M,loc} + \mathbb{E}h_{\theta_0}(X_1, X_1).$$

The distribution of $Z_{M,loc}$ is the limit distribution of $n\,\bar{U}_{n,n}$ defined in Proposition 5.2, where $\bar{U}_{n,n}$ denotes a U-statistic based on the underlying sample and whose kernel function is given by $h_n(x,y) = h^{(1)}(x,y) + n^{-1/2}\, h^{(2)}(x,y) + n^{-1/2}\, h^{(2)}(y,x) + n^{-1}\, h^{(3)}(x,y)$. Here, $h^{(1)} := h_{\theta_0}$ is defined as in Theorem 5.3,

$$\begin{aligned} h^{(2)}(x,y) := {} & [x_1 - g(x_2,\theta_0)]'[K(x_2,y_2,\theta_0)\, A(y_2) - V(x_2,\theta_0)T_a] \\ & + [l(x,\theta_0)]'[-V(y_2,\theta_0)\, A(y_2) + a(\theta_0)\, T_a], \end{aligned}$$

and

$$\begin{aligned} h^{(3)}(x,y) := {} & [A(x_2)]'\, K(x_2,y_2,\theta_0) A(y_2) - [A(x_2)]'\, V(x_2,\theta_0)\, T_a \\ & - [A(y_2)]'\, V(y_2,\theta_0)\, T_a + T_a'\, a(\theta_0)\, T_a \end{aligned}$$

with $x = (x_1', x_2')'$, $y = (y_1', y_2')'$, $x_1, y_1 \in \mathbb{R}^d$ and $x_2, y_2 \in \mathbb{R}^{dp+m\bar{p}}$. If additionally $\mathrm{var}(Z_M) > 0$, the following relations hold true:

$$\begin{aligned} & \liminf_{n\to\infty} \left[P(\widehat{T}^{(u)}_{n,n} > x) - P(\widehat{T}^{(u)}_{n} > x)\right] \geq 0 \quad \text{and} \\ & \liminf_{n\to\infty} \left[P(\widehat{T}^{(v)}_{n,n} > x) - P(\widehat{T}^{(v)}_{n} > x)\right] \geq 0 \quad \forall\, x \in \mathbb{R}. \end{aligned}$$

Proof. Invoking the identities $Y_{n,k} = Y_k + n^{-1/2}A(I_k)$ and $Y_{n,k} - g(I_k,\theta_0) = \epsilon_k + n^{-1/2}A(I_k)$, $k = 1,\dots,n$, we obtain $\widehat{T}^{(u)}_{n,n} - T^{(u)}_{n,n} = o_P(1)$ and $\widehat{T}^{(v)}_{n,n} - T^{(v)}_{n,n} = o_P(1)$ in analogy to the proof of Lemma 5.11 with

$$\begin{aligned} T^{(u)}_{n,n} := \frac{1}{n} \sum_{\substack{j,k=1 \\ j\neq k}}^{n} & [\epsilon_j + n^{-1/2}A(I_j)]'\, [K(I_j,I_k,\theta_0)\, (\epsilon_k + n^{-1/2}A(I_k)) - 2\, V(I_j,\theta_0) \\ & \times (l(X_k,\theta_0) + n^{-1/2}T_a)] + [l(X_j,\theta_0) + n^{-1/2}T_a]'\, a(\theta_0)\, [l(X_k,\theta_0) + n^{-1/2}T_a] \end{aligned}$$

and

$$T_{n,n}^{(v)} := \frac{1}{n} \sum_{j,k=1}^{n} [\epsilon_j + n^{-1/2} A(I_j)]' \, [K(I_j, I_k, \theta_0) \, (\epsilon_k + n^{-1/2} A(I_k)) - 2\, V(I_j, \theta_0)$$
$$\times (l(X_k, \theta_0) + n^{-1/2} T_a)] + [l(X_j, \theta_0) + n^{-1/2} T_a]' \, a(\theta_0) \, [l(X_k, \theta_0) + n^{-1/2} T_a],$$

respectively. To verify this relation for the V-statistics, we first prove that $\widehat{T}_{n,n}^{(v)} - \widetilde{T}_{n,n}^{(v)} = o_P(1)$, where

$$\widetilde{T}_{n,n}^{(v)} := \frac{1}{n} \sum_{j,k=1}^{n} \big\{ [\epsilon_j + n^{-1/2} \, A(I_j)]' \, K(I_j, I_k, \theta_0) \, [\epsilon_k + n^{-1/2} \, A(I_k)]$$
$$- 2[\epsilon_j + n^{-1/2} \, A(I_j)] K(I_j, I_k, \theta_0) \, g^{(1)}(I_k, \theta_0) \, (\widehat{\theta}_n - \theta_0)$$
$$+ (\widehat{\theta}_n - \theta_0)' \, [g^{(1)}(I_j, \theta_0)]' K(I_j, I_k, \theta_0) \, g^{(1)}(I_k, \theta_0) \, (\widehat{\theta}_n - \theta_0).$$

The difference $\widehat{T}_{n,n}^{(v)} - \widetilde{T}_{n,n}^{(v)}$ can be split up as before, one merely has to substitute ϵ_k by $\epsilon_k + n^{-1/2} A(I_k),\ k = 1, \ldots, n$, in the terms (5.32) up to (5.36). In accordance with the proof of Lemma 5.11, the counterpart of the term (5.32) can now be bounded by

$$o_P(1) + o_P(1) \sum_{i=1}^{d} n^{-3/2} \sum_{\substack{j,k=1 \\ j \neq k}}^{n} \Big[\epsilon_{j,i} \, k_i^{(1)}(I_j, I_k, \theta_0) \, A_i(I_k) + \epsilon_{j,i} \, k_i^{(2)}(I_j, I_k, \theta_0) \, A_i(I_k) \Big] .$$

Similarly to the investigation of the quantity (5.33) the latter sum is shown to be of order $O_P(1)$. According to the moment assumptions on the functions A, the counterparts of (5.33) up to (5.36) can be treated exactly as before. The verification of the relation $\widetilde{T}_{n,n}^{(v)} - T_{n,n}^{(v)} = o_P(1)$ can be carried out analogously to step 2 of the proof of Lemma 5.11. Here, the additional expression

$$\frac{1}{n^2} \sum_{j,k=1}^{n} [A(I_j)]' [K(I_j, I_k, \theta_0) \, g(I_k, \theta_0) - V(I_j, \theta_0)]$$

has to be analysed. The verification of its asymptotic negligibility is similar to the one of (5.33).

All considerations for the corresponding U-type statistics are similar and therefore omitted. To sum up, it remains to take the statistics $T_{n,n}^{(u)}$ and $T_{n,n}^{(v)}$ into further consideration. Proposition 5.2 yields

$$T_{n,n}^{(u)} \xrightarrow{d} Z_{M,loc} \quad \text{and} \quad T_{n,n}^{(v)} \xrightarrow{d} Z_{M,loc} + \mathbb{E} h_{\theta_0}(X_1, X_1).$$

Moreover, the additional assumptions regarding the functions $k_i,\ i = 1, \ldots, d$, assure that the kernel function of $T_{n,n}^{(u)}$ has the structure that is required in order to apply Lemma 5.6. Thus, we obtain

$$\liminf_{n\to\infty} \Big[P(T_{n,n}^{(u)} > x) - P(T_n^{(u)} > x) \Big] \geq 0 \quad \text{and}$$
$$\liminf_{n\to\infty} \Big[P(T_{n,n}^{(v)} > x) - P(T_n^{(v)} > x) \Big] \geq 0 \quad \forall\, x \in \mathbb{R}.$$

According to the approximations at the beginning of the proof, these inequalities finally yield those stated in the proposition. □

5.5.3 Bootstrapping critical values of the test

According to Theorem 5.3 the limit distributions of both test statistics are basically infinite weighted sums of products of correlated normal variables. The weights depend on the unknown parameter θ_0 in a complicated way. Thus, the (asymptotic) critical values of these tests can hardly be determined analytically or be tabulated. In a similar context, Escanciano [60] proposed to circumvent these difficulties by approximating the critical values with the aid of a wild bootstrap method. In order to prove consistency of this approach, he makes use of the additional assumption that the linearizing function l in (M2) has a product structure, i.e. $l(Y_k, I_k, \theta_0) = \epsilon_k \bar{l}(I_k, \theta_0)$ for some function $\bar{l}$. We suggest a parametric bootstrap algorithm instead, where l does not have to factorize as above. Recall that the random variables ε_k and Z_k, $k \in \mathbb{Z}$, are independent.

1. Determine $\widehat{\theta}_n$.
2. Calculate $\epsilon_k^{(n)} := Y_k - g(I_k, \widehat{\theta}_n)$ and $\bar{\epsilon} = n^{-1} \sum_{k=1}^n \varepsilon_k^{(n)}$.
3. Draw $\epsilon_1^*, \ldots, \epsilon_n^*$ via Efron's bootstrap from $(\epsilon_k^{(n)} - \bar{\epsilon})_k$.
4. Draw $Z_{1-\bar{p}}^*, \ldots, Z_n^*$ via Efron's bootstrap from $Z_{1-\bar{p}}, \ldots, Z_n$ (independently of $\epsilon_1^*, \ldots, \epsilon_n^*$).
5. Determine an initial vector $(Y_0^{*\prime}, \ldots, Y_{1-p}^{*\prime})'$ independently of $(\epsilon_k^*)_k$ and $(Z_k^*)_k$.
6. Generate the bootstrap sample $Y_k^* = g(I_k^*, \widehat{\theta}_n) + \epsilon_k^*$, where $I_k^* = (Y_{k-p}^{*\prime}, \ldots, Y_{k-1}^{*\prime}, Z_{k-\bar{p}+1}^{*\prime}, \ldots, Z_k^{*\prime})'$, $k = 1, \ldots, n$.
7. Compute the bootstrap parameter estimator $\widehat{\theta}_n^*$.
8. Calculate the bootstrap versions of the test statistics
$$\widehat{T}_n^{(u)*} := \frac{1}{n} \sum_{\substack{j,k=1 \\ j \neq k}}^{n} [Y_j^* - g(I_j^*, \widehat{\theta}_n^*)]' \, K(I_j^*, I_k^*, \widehat{\theta}_n^*) \, [Y_k^* - g(I_k^*, \widehat{\theta}_n^*)],$$
$$\widehat{T}_n^{(v)*} := \frac{1}{n} \sum_{j,k=1}^{n} [Y_j^* - g(I_j^*, \widehat{\theta}_n^*)]' \, K(I_j^*, I_k^*, \widehat{\theta}_n^*) \, [Y_k^* - g(I_k^*, \widehat{\theta}_n^*)].$$
9. Define the critical values $t_\alpha^{(u)*}$ and $t_\alpha^{(v)*}$ as the $(1-\alpha)$-quantiles of the (conditional) distributions of $\widehat{T}_n^{(u)*}$ and $\widehat{T}_n^{(v)*}$, respectively.
Reject $\mathcal{H}_0$ if $\widehat{T}_n^{(u)} > t_\alpha^{(u)*}$ and $\widehat{T}_n^{(v)} > t_\alpha^{(v)*}$, respectively.

Validity of the bootstrap algorithm can be verified if we assume besides (M1) to (M3):

(M5) **(i)** The function g satisfies $P(\mathbb{E}^* \| g(x^*, Z_1^*, \widehat{\theta}_n) \|_2 \leq K) \longrightarrow_{n\to\infty} 1$ for some $x^* \in \mathbb{R}^{dp+m\bar{p}}$ and a $K < \infty$. It admits the expansion (5.28) with $\theta = \widehat{\theta}_n$ and continuous functions $(H_j(\cdot, \widehat{\theta}_n))_{j=1}^{P+1}$ with $P(\sum_{j=1}^P \mathbb{E}^* |H_j(Z_1^*, \widehat{\theta}_n)| \leq 1 - \delta) \longrightarrow_{n\to\infty} 1$

for some $\delta \in (0,1)$. Moreover, the moments

$$\mathbb{E}^* \max_{a\in[-A,A]^{dp+m\bar{p}}} |H(I_1^*+a, I_1^*, \widehat{\theta}_n)|^{4+\varepsilon},\ \mathbb{E}^* \max_{a\in[-A,A]^{dp+m\bar{p}}} |H(I_1^*, I_1^*+a, \widehat{\theta}_n)|^{4+\varepsilon},$$
$$\mathbb{E}^* \max_{a\in[-A,A]^{dp+m\bar{p}}} |H(I_1^*+a, \widetilde{I}_1^*, \widehat{\theta}_n)|^{4+\varepsilon},\ \mathbb{E}^* \max_{a\in[-A,A]^{dp+m\bar{p}}} |H(I_1^*, \widetilde{I}_1^*+a, \widehat{\theta}_n)|^{4+\varepsilon}$$

of the function $H := \sum_{j=1}^{p+1} H_j$ are of order $O_P(1)$ for some $\varepsilon, A > 0$ and an independent copy $\widetilde{I}_1^*$ of I_1^*, conditionally on $X_1, \ldots, X_n$.

(ii) The function $g(x,\cdot)$ is three times continuously differentiable for all $x \in \mathbb{R}^{dp+m\bar{p}}$. The components of the first two partial derivatives of g w.r.t. θ satisfy $\sum_{i=1}^d \sum_{\alpha,\beta=1}^q \mathbb{E}^*[|g_i(I_1^*, \widehat{\theta}_n)| + |g_{i;\alpha}^{(1)}(I_1^*, \widehat{\theta}_n)| + |g_{i;\alpha,\beta}^{(2)}(I_1^*, \widehat{\theta}_n)|]^{4+\varepsilon} = O_P(1)$. Moreover, (5.29) holds true with $\theta = \widehat{\theta}_n$, $\mathbb{E}|f_g(I_1^*, \widetilde{I}_1^*, \widehat{\theta}_n)|^{4+\varepsilon} = O_P(1)$, and $P(\sup_{j,k\in\mathbb{N}} \mathbb{E}^*|f_g(I_j^*, I_k^*, \widehat{\theta}_n)|^{4+\varepsilon} \le K) \longrightarrow_{n\to\infty} 1$. The function M, defined in (M1), satisfies $P(\mathbb{E}^*[M(I_1^*)]^2 \le K) \longrightarrow_{n\to\infty} 1$.

(iii) The function g fulfils $\|g(y,\theta_1) - g(y,\theta_2)\|_1 \le L(y)\,\|\theta_1 - \theta_2\|_1$ where $L : \mathbb{R}^{dp+m\bar{p}} \to \mathbb{R}_+$ is a locally Lipschitz continuous function with $\mathbb{E}\|L(I_1)\|_1^{4+\varepsilon} < \infty$ for some $\varepsilon > 0$. The random vector $Y_1 - g(I_1, \theta_0)$ has continuous marginal distribution functions.

The bootstrap estimator is assumed to satisfy:

(M6) **(i)** The sequence of estimators $\widehat{\theta}_n^*$ admits the expansion

$$\widehat{\theta}_n^* - \widehat{\theta}_n = \frac{1}{n}\sum_{k=1}^n l(X_k^*, \widehat{\theta}_n) + o_{P^*}(n^{-1/2}) \tag{5.38}$$

with $X_k^* = (Y_k^{*\prime}, I_k^{*\prime})'$ and $\mathbb{E}^* l(X_1^*, \widehat{\theta}_n) = 0$, $\mathbb{E}^*\|l(X_1^*, \widehat{\theta}_n)\|_1^{4+\varepsilon} = O_P(1)$.

(ii) Moreover, $\|l(x,\widehat{\theta}_n) - l(\bar{x},\widehat{\theta}_n)\|_1 \le f_l(x,\bar{x},\widehat{\theta}_n)\|x-\bar{x}\|_1$, where f_l is a symmetric and continuous function with

$$P\left(\sup_{j,k\in\mathbb{N}} \mathbb{E}^* \max_{a\in[-A,A]^{pd+\bar{p}m}} |f_l(X_j^*+a, X_k^*, \widehat{\theta}_n)|^{4+\varepsilon} \le K\right) \underset{\mathrm{n}\to\infty}{\longrightarrow} 1$$

and $\mathbb{E}^* \max_{a\in[-A,A]^{pd+\bar{p}m}} |f_l(X_1^*+a, \widetilde{X}_1^*, \widehat{\theta}_n)|^{4+\varepsilon} = O_P(1)$ for any independent copy $\widetilde{X}_1^*$ of X_1^*, some $\varepsilon, A > 0$, and a $K < \infty$.

The assumptions (M5)(i),(ii) and (M6) can be interpreted as the bootstrap counterparts to (M1) and (M2). Note that the condition (M5) holds for instance when the function g is linear with coefficients whose absolute values sum up to some constant that is less than one and if the corresponding parameter estimators converge. The additional assumption (M5)(iii) is required to prove that the fidis of the bootstrap process converge to those of the original process.

Applying the results of Chapter 4, we obtain bootstrap consistency for our test statistics.

Proposition 5.8. *Let the assumptions of Lemma 5.11 hold true. Suppose that the conditions* $(M5)$ *and* $(M6)$ *are fulfilled and that* $\widehat{T}_n^{(u)*}$ *as well as* $\widehat{T}_n^{(v)*}$ *are generated via the aforementioned algorithm, where the initial response vector* Y_0^* *is drawn from the stationary bootstrap distribution if it exists. Then, under* $\mathcal{H}_0$,

$$\widehat{T}_n^{(u)*} \xrightarrow{d} Z_M \quad \text{and}$$

$$\widehat{T}_n^{(v)*} \xrightarrow{d} Z_M + \mathbb{E}\left\{\epsilon_1' \left[K(I_1, I_1, \theta_0)\,\epsilon_1 - 2V(I_1, \theta_0)\, l(X_1, \theta_0)\right] + l(X_1, \theta_0)'\, a(\theta_0)\, l(X_1, \theta_0)\right\},$$

in probability as $n \to \infty$, *where* Z_M *is defined as in Theorem 5.3. Moreover, if* $\operatorname{var}(Z_M) > 0$,

$$\sup_{-\infty<x<\infty} \left| P(\widehat{T}_n^{(u)*} \le x | X_1, ..., X_n) - P(\widehat{T}_n^{(u)} \le x) \right| \xrightarrow{P} 0,$$
$$\sup_{-\infty<x<\infty} \left| P(\widehat{T}_n^{(v)*} \le x | X_1, ..., X_n) - P(\widehat{T}_n^{(v)} \le x) \right| \xrightarrow{P} 0.$$

Remark 5.6. The result holds in an idealized case, i.e. when the initial bootstrap vector Y_0^* has the possibly unknown stationary bootstrap distribution provided its existence. However, we conjecture that it remains valid for any "reasonable" starting vector, cf. Remark 4.6.

Proof. The proof is carried out in two steps. First, we verify the bootstrap sample $(X_k^*)_k$ to satisfy the condition (A1*) of Chapter 4. Afterwards, the assertions of the theorem is proved.

Step 1: Verification of (A1*).

We apply our results of Subsection 4.4.4. To this end, the notation

$$X_t = \begin{pmatrix} Y_t \\ Z_t \end{pmatrix} = \begin{pmatrix} g(Y_{t-p}, \ldots, Y_{t-1}, Z_{t-\bar{p}+1}, \ldots, Z_t, \theta_0) + \epsilon_t \\ Z_t \end{pmatrix} =: G(X_{t-1}, \ldots, X_{t-P}, \varepsilon_t; \theta_0)$$

is introduced with $\varepsilon_t = (\epsilon_t', Z_t')'$. Thus, our bootstrap procedure is equivalent to the Algorithm [B3] if $\varepsilon_1^* := (\epsilon_t^{*\prime}, Z_t^{*\prime})' \xrightarrow{d} \varepsilon_1$, in probability. The validity of (A1*) then follows from Lemma 4.8 if its prerequisites are fulfilled. Under (M1) the process $(X_t)_t$ satisfies the conditions (I) and (II) of Subsection 4.4.4. Moreover, the prerequisites (a) and (b) of Lemma 4.8 are satisfied due to (M5)(i). Therefore, in order to obtain (A1*), it remains to verify $\varepsilon_1^* \xrightarrow{d} \varepsilon_1$, in probability. First, note that $Z_1^* \xrightarrow{d} Z_1$, in probability, since the variables Z_k^* are drawn via Efron's bootstrap from $Z_{1-\bar{p}}, \ldots, Z_n$. Next, we prove $\epsilon_1^* \xrightarrow{d} \epsilon_1$, in probability, more precisely, we show that $P(|F_{\epsilon_1^*}(x) - F_{\epsilon_1}(x)| > \eta) \le \delta$ for any $\eta, \delta > 0$ and all $n \ge n_0(\eta, \delta)$ and arbitrary fixed $x \in \mathbb{R}^d$. To this end, one splits up

$$P(|F_{\epsilon_1^*}(x) - F_{\epsilon_1}(x)| > \eta) \le P\Big(\sum_{r=1}^d \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{x_r - \|g(I_j, \widehat{\theta}_n) - g(I_j, \theta_0)\|_1 - \|n^{-1}\sum_{i=1}^n [Y_i - g(I_i, \widehat{\theta}_n)]\|_1 \le \epsilon_{j,r} \le x_r + \|g(I_j, \widehat{\theta}_n) + g(I_j, \theta_0)\|_1 + \|n^{-1}\sum_{i=1}^n [Y_i - g(I_i, \widehat{\theta}_n)]\|_1} > \frac{\eta}{2}\Big)$$
$$+ P\left(\left|\frac{1}{n}\sum_{j=1}^n \mathbb{1}_{\epsilon_j \preceq x} - F_{\epsilon_1}(x)\right| > \frac{\eta}{2}\right).$$

The first summand of the r.h.s. can be bounded by employing Markov's inequality:

$$\begin{aligned}
&\frac{C}{\eta}\Big[\sum_{r=1}^{d} P\big(\epsilon_{1,r}\in[x_r-3\kappa,x_r+3\kappa]\big)+P\Big(\sum_{r=1}^{d}\Big|\frac{1}{n}\sum_{i=1}^{n}\epsilon_{i,r}\Big|>\kappa\Big)\\
&\quad+P\Big(\frac{1}{n}\sum_{i=1}^{n}\Big\|g(I_i,\widehat{\theta}_n)-g(I_i,\theta)\Big\|_1>\kappa\Big)+\frac{1}{n}\sum_{j=1}^{n}P\Big(\Big\|g(I_j,\widehat{\theta}_n)-g(I_j,\theta)\Big\|_1>\kappa\Big)\Big]\\
&\le\frac{C}{\eta}\Big[\sum_{r=1}^{d} P\big(\epsilon_{1,r}\in[x_r-3\kappa,x_r+3\kappa]\big)+n^{-1}\kappa^{-2}+P(L(I_1)>1/\kappa)\\
&\quad+P\Big(\|\widehat{\theta}_n-\theta_0\|_1>\kappa^2\Big)+P\Big(n^{-1}\sum_{i=1}^{n}L(I_i)>1/\kappa\Big)\Big],
\end{aligned}$$

which is less than $\delta/2$ for sufficiently small $\kappa>0$ and all $n\ge n_0$ for some $n_0(\eta,\delta,\kappa)$. Concerning the second summand, Chebychev's inequality yields

$$\begin{aligned}
&P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_{\epsilon_j\preceq x}-F_{\epsilon_1}(x)\right|>\eta\right)\\
&\le\frac{1}{n\,\eta^2}+\frac{2}{(n\,\eta)^2}\sum_{1\le j<k\le n}\mathbb{E}\big\{[\mathbb{1}_{\epsilon_j\preceq x}-F_{\epsilon_1}(x)][\mathbb{1}_{\epsilon_k\preceq x}-F_{\epsilon_1}(x)]\big\}\\
&\le\frac{C}{n\,\eta^2}.
\end{aligned}$$

due to the independence of the innovations. Hence, for all $\delta>0$ and any sufficiently small $\kappa>0$ there exists an $n_0(\kappa,\delta)$ with $P(|F_{\epsilon_1^*}(x)-F_{\epsilon_1}(x)|>\eta)<\delta,\ \forall\,n\ge n_0$. This completes the proof of the convergence of the (conditional) distribution of the bootstrap innovations to P_{ϵ_1}. Due to the independence of Z_1 and ϵ_1 as well as of Z_1^* and ϵ_1^*, we have $\varepsilon_1^*\xrightarrow{d}\varepsilon_1$, in probability. Therefore, all prerequisites of Lemma 4.8 are fulfilled which assures the validity of (A1*).

<u>*Step 2:*</u> *Convergence of $\widehat{T}_n^{(u)*}$ and $\widehat{T}_n^{(v)*}$.*

First, we have $\mathbb{E}^*|Z_k^*|^2\xrightarrow{P}\mathbb{E}|Z_1|^2$. Moreover, $\mathbb{E}^*|Y_k^*|^{4+\varepsilon}=O_P(1)$ for some $\varepsilon>0$ as, on the one hand, $\mathbb{E}^*|g(I_k^*,\widehat{\theta}_n)|^{4+\varepsilon}=O_P(1)$ by assumption and, on the other hand,

$$\begin{aligned}
\mathbb{E}^*|\epsilon_1^*|^{4+\varepsilon}&\le C\left[\mathbb{E}|\epsilon_1|^{4+\varepsilon}+\frac{1}{n}\sum_{k=1}^{n}|g(I_k,\widehat{\theta}_n)-g(I_k,\theta_0)|^{4+\varepsilon}\right]\\
&\le C+o_P(1)\,\frac{1}{n}\sum_{k=1}^{n}|L(I_k)|^{4+\varepsilon}\\
&=O_P(1)
\end{aligned}$$

in view of Lemma 2.3. W.l.o.g. we consider $\widehat{T}_n^{(v)*}$ only. Note that for all $\alpha,\beta\in\{1,\dots,q\}$

$$|\mathbb{E}^*([g^{(1)}(I_1^*,\widehat{\theta}_n)'K(I_1^*,\widetilde{I}_1^*,\widehat{\theta}_n)g^{(1)}(\widetilde{I}_1^*,\widehat{\theta}_n)]_{\alpha,\beta})-a(\theta_0)|\xrightarrow{P}0,\ \forall\,\alpha,\beta\in\{1,\dots,q\},$$

where $\widetilde{I}_1^*$ denotes and independent copy of I_1^*, conditionally on $\mathbb{X}_n$. Following the lines of the proof of Lemma 5.11, one obtains $\widehat{T}_n^{(v)*} - T_n^{(v)*} = o_{P^*}(1)$ with

$$T_n^{(v)*} := \frac{1}{n} \sum_{j,k=1}^{n} \big\{ \epsilon_j^{*\prime} \big[K(I_j^*, I_k^*, \widehat{\theta}_n) \epsilon_k^* - 2\, V(I_j^*, \theta_0)\, l(X_k^*, \widehat{\theta}_n) \big] \\ + [l(X_j^*, \widehat{\theta}_n)]'\, a(\theta_0)\, l(X_k^*, \widehat{\theta}_n) \big\}$$

if for all $\alpha, \beta \in \{1, \ldots, q\}$,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \epsilon_j^{*\prime} \, [\mathbb{E}^*(k_\alpha(I_j^*, \widetilde{I}_j^*, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(\widetilde{I}_j^*, \widehat{\theta}_n) | I_j^*) - V_{\alpha,\beta}(I_j^*, \theta_0)] = o_{P^*}(1). \tag{5.39}$$

In order to verify this relation, we prove asymptotic negligibility of

$$\mathbb{E}^* \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \epsilon_j^{*\prime} \left[\int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(I_j^*, x, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n} - V_{\alpha,\beta}(I_j^*, \theta_0) \right] \right\}^2 .$$

The introduction of a copy $\widetilde{X}_k^*$ of $X_k^*,\ k \in \mathbb{N}$, that is independent of X_j^*, conditionally on $X_1, \ldots, X_n$, and such that the bootstrap counterpart of inequality (5.31) holds for $n = k - j > P + 1$ with probability tending to one, leads to

$$\begin{aligned}
&\mathbb{E}^* \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \epsilon_j^{*\prime} \left[\int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(I_j^*, x, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_{\alpha,\beta}(I_j^*, \widehat{\theta}_n) \right] \right\}^2 \\
&\le O_P(1) \left(\mathbb{E}^* \left| \int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(I_1^*, x, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_{\alpha,\beta}(I_1^*, \theta_0) \right|^2 \right)^{1/2} \\
&\quad + \frac{2}{n} \sum_{\substack{j,k=1 \\ k-j>P}}^{n} \mathbb{E}^* \Big(\epsilon_j^{*\prime} \Big[\int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(I_j^*, x, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_{\alpha,\beta}(I_j^*, \theta_0) \Big] \\
&\quad \times \Big\{ \epsilon_k^{*\prime} \Big[\int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(I_k^*, x, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n} - V_{\alpha,\beta}(I_k^*, \theta_0) \Big] \\
&\quad - \widetilde{\epsilon}_k^{*\prime} \Big[\int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(\widetilde{I}_k^*, x, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_{\alpha,\beta}(\widetilde{I}_k^*, \theta_0) \Big] \Big\} \Big) \\
&\le O_P(1) \left(\mathbb{E}^* \left| \int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(I_1^*, x, \widehat{\theta}_n)\, g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_{\alpha,\beta}(I_1^*, \theta_0) \right|^{1/\delta} \right)^{\delta}
\end{aligned}$$

for some $\delta \in (0, 1/2]$. It remains to verify that the quantity in round brackets can be bounded with probability tending to one by any given $\varepsilon > 0$ if n is sufficiently large. For this purpose we introduce a compact set D such that

$$\mathbb{E}^* \left| \int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(I_j^*, x, \widehat{\theta}_n) g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_{\alpha,\beta}(I_j^*, \theta_0) \right|^{1/\delta} \mathbb{1}_{I_1^* \notin D} \le C P(I_1^* \notin D) \le \frac{\varepsilon}{2}$$

holds with probability tending to one. Moreover, according to the convergence of the fidis and due to Lipschitz continuity of the elements of K, we have

$$\max_{z \in D} \left| \int_{\mathbb{R}^{dp+m\bar{p}}} k_\alpha(z, x, \widehat{\theta}_n) g_{\alpha,\beta}^{(1)}(x, \widehat{\theta}_n) P_{I_1^*|\mathbb{X}_n}(dx) - V_\alpha(z, \theta_0) \right|^{1/\delta} \le \frac{\varepsilon}{2}$$

with probability tending to one, which finally yields (5.39).

Thus, it suffices to show the desired convergence for $T_n^{(v)*}$ instead of $\widehat{T}_n^{(v)*}$. The statistic $T_n^{(v)*}$ is the bootstrap counterpart of the degenerate V-statistic $T_n^{(v)}$ of Lemma 5.11. Define h and f_θ as in the proof of Theorem 5.3, we get

$$P\left(\sup_{k\in\mathbb{N}} \mathbb{E}^*|h(X_1^*, X_{1+k}^*, \widehat{\theta}_n)|^{2+\varepsilon} + \mathbb{E}^*|h(X_1^*, \widetilde{X}_1^*, \widehat{\theta}_n)|^{2+\varepsilon} \leq K\right) \xrightarrow{P} 1$$

as well as

$$P\left(\sup_{k_1,\ldots,k_5\in\mathbb{N}} \mathbb{E}^*\left(\max_{a_1,a_2\in[-A,A]^{d(p+1)+m\bar{p}}} [f_{\widehat{\theta}_n}(\bar{X}_{k_1}^*, \bar{X}_{k_2}^* + a_1, \bar{X}_{k_3}^*, \bar{X}_{k_4}^* + a_2)]^\eta \|\bar{X}_{k_5}^*\|_1\right) \leq K\right) \xrightarrow{P} 1$$

for some $\varepsilon > 0$, $K < \infty$ and any $(\bar{X}_{k_1}^{*\prime}, \ldots, \bar{X}_{k_5}^{*\prime})'$ consisting of independent subvectors $(\bar{X}_{k_{j_1(m)}}^{*\prime}, \ldots, \bar{X}_{k_{j_l(m)}}^{*\prime})' \stackrel{d}{=} (X_{k_{j_1(m)}}^{*\prime}, \ldots, X_{k_{j_l(m)}}^{*\prime})'$, $l, m = 1, \ldots, 5$. Hence, the conditions (A2*) and (A3*) are fulfilled. Moreover, according to the first step of the proof, the bootstrap sample satisfies (A1*). Consequently, we obtain bootstrap consistency by Theorem 4.1. □

In accordance with both of the foregoing test procedures, this bootstrap-based test has asymptotically the correct size under the null hypothesis, is consistent against fixed alternatives, and is asymptotically unbiased against certain Pitman alternatives. Again we conjecture that the proposed test has non-trivial power against a subclass of these local alternatives. Reasons for this are that, on the one hand, the bootstrap algorithm imitates the null under $\mathcal{H}_{1,n}$, cf. the proof of Corollary 5.9 below and on the other hand, $Z_{M,loc} = D_a + Z_M$, $D_a > 0$, can be obtained for a subclass of Pitman alternatives in the case of univariate response variables according to Escanciano [60], under some additional assumptions on the kernel and on the sequence of parameter estimators.

Corollary 5.3. *Suppose that the conditions (M5) and (M6) are fulfilled and let $\alpha \in (0,1)$. Moreover, assume that the prerequisites of Lemma 5.11 and Lemma 5.12 hold. The proposed bootstrap test based on the algorithm above satisfies*

$$\lim_{n\to\infty} P(\widehat{T}_n^{(u)} > t_\alpha^{(u)*}) = \begin{cases} \alpha & \text{if } \mathcal{H}_0 \text{ is true and } \operatorname{var}(Z_M) > 0, \\ 1 & \text{if } \mathcal{H}_1 \text{ is true} \end{cases}$$

and

$$\lim_{n\to\infty} P(\widehat{T}_n^{(v)} > t_\alpha^{(v)*}) = \begin{cases} \alpha & \text{if } \mathcal{H}_0 \text{ is true and } \operatorname{var}(Z_M) > 0, \\ 1 & \text{if } \mathcal{H}_1 \text{ is true,} \end{cases}$$

i.e. the bootstrap-based test is consistent. Under $\mathcal{H}_{1,n}$ and the additional assumptions of Proposition 5.7,

$$\liminf_{n\to\infty} P(\widehat{T}_{n,n}^{(u)} > t_\alpha^{(u)*}) \geq \alpha \quad \text{and} \quad \liminf_{n\to\infty} P(\widehat{T}_{n,n}^{(v)} > t_\alpha^{(v)*}) \geq \alpha.$$

Proof. Step 1: *Behaviour under* $\mathcal{H}_0$.
These results follow from Proposition 5.8.
Step 2: *Behaviour under* $\mathcal{H}_1$.
Under our assumptions one can prove that the proposed bootstrap method imitates a null situation, namely the (unique) stationary solution of $\bar{Y}_k = g(\bar{Y}_{k-p},\ldots,\bar{Y}_{k-1},Z_{k-\bar{p}+1},\ldots,Z_k,\theta_0)+\bar{\epsilon}_k$. Here, $((Z_k',\bar{\epsilon}_k')')_k$ is a sequence of i.i.d. random variables, where Z_1 and $\bar{\epsilon}_1$ are independent and $\bar{\epsilon}_1 \stackrel{d}{=} Y_1 - g(I_1,\theta_0) - \mathbb{E}Y_1 + \mathbb{E}g(I_1,\theta_0)$. To this end, the arguments of step 1 of the proof of Proposition 5.8 are be invoked. Thus, it remains to show that $\epsilon_1^* \xrightarrow{d} \bar{\epsilon}_1$, in probability. We have

$$
\begin{aligned}
&P(|F_{\epsilon_1^*}(x) - F_{\epsilon_1}(x)| > \eta)\\
&\le P\Big(\sum_{r=1}^d \frac{1}{n}\sum_{j=1}^n \mathbb{1}_{x_r - \|g(I_j,\widehat{\theta}_n)-g(I_j,\theta_0)\|_1 - \|n^{-1}\sum_{i=1}^n [Y_i - g(I_i,\widehat{\theta}_n) - \mathbb{E}Y_1 + \mathbb{E}g(I_1,\theta_0)]\|_1 \le \bar{\epsilon}_{j,r}}\\
&\qquad {}_{\le x_r + \|g(I_j,\widehat{\theta}_n)+g(I_j,\theta_0)\|_1 + \|n^{-1}\sum_{i=1}^n [Y_i - g(I_i,\widehat{\theta}_n) - \mathbb{E}Y_1 + \mathbb{E}g(I_1,\theta_0)]\|_1} > \frac{\eta}{2}\Big)\\
&\quad + P\left(\left|\frac{1}{n}\sum_{j=1}^n \mathbb{1}_{\bar{\epsilon}_j \preceq x} - F_{\bar{\epsilon}_1}(x)\right| > \frac{\eta}{2}\right).
\end{aligned}
$$

An upper bound for the first summand on the r.h.s. is obtained similarly as in the proof of Proposition 5.8. However, the approximation of the second summand has to be modified since the underlying variables are no longer independent here. Thus, we have to use certain coupling arguments to very asymptotic negligibility of

$$
\begin{aligned}
&P\left(\left|\frac{1}{n}\sum_{j=1}^n \mathbb{1}_{\bar{\epsilon}_j \preceq x} - F_{\bar{\epsilon}_1}(x)\right| > \eta\right)\\
&\le \frac{1}{n\,\eta^2} + \frac{2}{(n\,\eta)^2}\sum_{1\le j<k\le n} \mathbb{E}\big\{[\mathbb{1}_{\bar{\epsilon}_j \preceq x} - F_{\bar{\epsilon}_1}(x)][\mathbb{1}_{\bar{\epsilon}_k \preceq x} - F_{\bar{\epsilon}_1}(x)]\big\}.
\end{aligned}
$$

Unfortunately, the involved terms are not continuous and thus the coupling techniques cannot be invoked directly. Therefore, the introduction of a smoothing function is helpful in order to approximate the expectations above. For $a > 0$ we define

$$
G_a(z) := \int_{\mathbb{R}^d} w_a(u)\,\mathbb{1}_{z \preceq x+u}\,du - F_{\bar{\epsilon}_1}(x), \quad z \in \mathbb{R}^d.
$$

Here, $(w_a)_{a>0}$ is a family of nonnegative functions with $\operatorname{supp}(w_a) \subseteq \{u \in \mathbb{R}^d \mid u \succeq 0, \|u\|_1 \le a\}$, $\int_{\mathbb{R}^d} w_a(u)\,du = 1$ and $\|w_a\|_\infty \le Ca^{-d}$. Thus, G_a is Lipschitz continuous with $\operatorname{Lip}(G_a) \le Ca^{-1}$ as

$$
|G_a(z) - G_a(\bar{z})| \le \sum_{i=1}^d \int_{\mathbb{R}^d} w_a(u)\,[\mathbb{1}_{z_i - x_i \le u_i \le \bar{z}_i - x_i} + \mathbb{1}_{\bar{z}_i - x_i \le u_i \le z_i - x_i}]\,du \le Ca^{-1}\|z - \bar{z}\|_1.
$$

Let $\widetilde{\epsilon}_k$ be a copy of $\bar{\epsilon}_k$ that is independent of $\bar{\epsilon}_j$ such that (5.31) holds with $n = k - j$ for

$k-j>P+1$. Then,

$$\begin{aligned}
&\frac{2}{(n\,\eta)^2}\sum_{1\le j<k\le n}\mathbb{E}\big\{[\mathbb{1}_{\bar\epsilon_j\preceq x}-F_{\bar\epsilon_1}(x)][\mathbb{1}_{\bar\epsilon_k\preceq x}-F_{\bar\epsilon_1}(x)]\big\}\\
&=\frac{2}{(n\,\eta)^2}\sum_{1\le j<k\le n}\mathbb{E}\Big\{[\mathbb{1}_{\bar\epsilon_j\preceq x}-F_{\bar\epsilon_1}(x)]\Big[\mathbb{1}_{\bar\epsilon_k\preceq x}-\int_{\mathbb{R}^d}w_a(u)\mathbb{1}_{\bar\epsilon_k\preceq x+u}du\Big]\Big\}\\
&\quad+\frac{2}{(n\,\eta)^2}\sum_{1\le j<k\le n}\mathbb{E}\big\{[\mathbb{1}_{\bar\epsilon_j\preceq x}-F_{\bar\epsilon_1}(x)]G_a(\bar\epsilon_k)]\big\}\\
&\le\frac{2}{\eta^2}\sum_{r=1}^d P(x_r\le\bar\epsilon_{1,r}\le x_r+a)+\frac{C}{\eta^2\,n}+\frac{2}{(n\,\eta)^2}\sum_{\substack{1\le j<k\le n\\ k-j>P+1}}\mathbb{E}|G_a(\bar\epsilon_k)-G_a(\widetilde\epsilon_k)|\\
&\le\frac{2}{\eta^2}\sum_{r=1}^d P(x_r\le\bar\epsilon_{1,r}\le x_r+a)+\frac{C}{\eta^2\,n}+\frac{C}{\eta^2\,a\,n}\sum_{k=1}^\infty\rho^k
\end{aligned}$$

for some $\rho\in(0,1)$. Thus, the latter expression can be bounded by any $\delta>0$ whenever a is sufficiently small and $n\ge n_0(\delta,a)$. This finally implies the validity of (A1*) in this case. Step 2 of the proof of Proposition 5.8 remains valid under $\mathcal{H}_1$. Consequently, the quantiles of the bootstrap statistics are bounded with probability tending to one. In accordance with Lemma 5.11, we obtain the desired assertions.

Step 3: Behaviour under $\mathcal{H}_{1,n}$.

Here, the bootstrap algorithm imitates the case $A(I_t)=0$, which can be proved by following the lines of step 1 of the proof of Proposition 5.8 and applying the inequality

$$\begin{aligned}
&P(|F_{\epsilon_1^*}(x)-F_{\epsilon_1}(x)|>\eta)\\
&\le P\Big(\sum_{r=1}^d\frac{1}{n}\sum_{j=1}^n\mathbb{1}_{x_r-n^{-1/2}\|A(I_j)\|_1-\|g(I_j,\widehat\theta_n)-g(I_j,\theta_0)\|_1-\|n^{-1}\sum_{i=1}^n[Y_i-g(I_i,\widehat\theta_n)]\|_1\le\epsilon_{j,r}}\\
&\qquad{}_{\le x_r+n^{-1/2}\|A(I_j)\|_1+\|g(I_j,\widehat\theta_n)+g(I_j,\theta_0)\|_1+\|n^{-1}\sum_{i=1}^n[Y_i-g(I_i,\widehat\theta_n)]\|_1}>\frac{\eta}{2}\Big)\\
&\quad+P\left(\left|\frac{1}{n}\sum_{j=1}^n\mathbb{1}_{\epsilon_j\preceq x}-F_{\epsilon_1}(x)\right|>\frac{\eta}{2}\right).
\end{aligned}$$

To this end, note that, in analogy to $\mathcal{H}_0$, we have $n^{-1}\sum_{i=1}^n[Y_i-g(I_i,\theta_0)]\xrightarrow{P}0_d$ and, additionally, $P(n^{-1/2}\|A(I_1)\|_1>\kappa)\longrightarrow_{n\to\infty}0$ for all $\kappa>0$. Similar to the proof of Corollary 5.1, the assertion now follows from Proposition 5.7. $\square$

6 Summary and future perspectives

To sum up, two major results of the thesis can be highlighted. First, the limit distributions of degenerate degree-2 U- and V-type statistics have been derived under assumptions that are easy check, namely

- stationarity and τ-dependence of the underlying observations,
- smoothness conditions and moment constraints concerning the kernel function.

In contrast, most of the results available in the literature have been obtained under assumptions that are difficult or even impossible to check in many cases. The only competitor is the work of Babbel [6] that is concerned with U-statistics of mixing random variables. However, this result is hardly applicable for proving consistency of model-based bootstrap methods since the bootstrap counterparts of mixing process can usually not be proved to satisfy any mixing condition. Moreover, difficulties occur when one intends to construct hypothesis tests as those illuminated in Chapter 5 on the basis of Babbel's [6] findings. Her assumption $\iint_{(0,1)\times(0,1)} h(x,y)\, dP^{X_s,X_{s+m}}(x,y) = 0,\ \forall s \in \mathbb{Z},\ m \in \mathbb{N}$, turns out to be too restrictive.

Invoking the concept of τ-dependence, we have succeeded in providing consistency results for the bootstrap versions of the statistics under consideration. The crucial assumptions can be summarized as follows:

- stationarity and τ-dependence of the bootstrap variables, with probability tending to one,
- convergence of the two-dimensional distributions of the bootstrap process towards the respective two-dimensional distributions of the original process, in probability,
- moment constraints concerning the kernel, and
- a positive variance of the limit distribution of the U-statistics.

We have illustrated how both results can be applied in the field of hypothesis testing. The tests considered in Section 5.3 and Section 5.4 apply to certain properties of the marginal distribution of a time series. So far, model-based bootstrap methods were invoked to approximate critical values of these tests. Those methods are used when the observed process is assumed to arise from a certain time series model. However, those model assumptions often characterize the conditional rather than the unconditional distribution of the underlying random variables. From this point of view, it is desirable to derive consistency of block bootstrap procedures for degenerate U- and V-statistics, too. While in the non-degenerate case these methods have been investigated by Dehling and Wendler [42], there are no such results in the degenerate case so far. This gap shall be bridged in future work.

Moreover, all three tests of Chapter 5 are concerned with testing composite hypotheses. Hence, the test statistics involve parameter estimators and are not degenerate themselves but can be approximated by degenerate statistics of U- and V-type, respectively. To circumvent these case-by-case approximations, it might be of interest to extend the theory of Chapter 3 to statistics with estimated parameters that are degenerate in the limit. For U- and V-statistics of i.i.d. observations, those problems have already been considered by de Wet and Randles [48].

Another possible extension of the results that have been achieved in the present thesis constitutes the consideration of two-sample statistics. More precisely, one can focus on

$$U_{n_1,n_2} = \frac{1}{n_1(n_1-1)n_2(n_2-1)} \sum_{\substack{i,j=1\\ i\neq j}}^{n_1} \sum_{\substack{k,l=1\\ k\neq l}}^{n_2} h\left(X_i^{(1)}, X_j^{(1)}; X_k^{(2)}, X_l^{(2)}\right)$$

and

$$V_{n_1,n_2} = \frac{1}{n_1^2 n_2^2} \sum_{i,j=1}^{n_1} \sum_{k,l=1}^{n_2} h\left(X_i^{(1)}, X_j^{(1)}; X_k^{(2)}, X_l^{(2)}\right)$$

given two samples $X_1^{(1)},\ldots,X_{n_1}^{(1)}$ and $X_1^{(2)},\ldots,X_{n_2}^{(2)}$. The asymptotic theory of these statistics is well developed in the i.i.d. setting, see e.g. Koroljuk and Borovskich [80], Section 4.5. It can be applied to test the hypothesis $\mathcal{H}_0 : P_{X_1^{(1)}} = P_{X_1^{(2)}}$ against its alternative $\mathcal{H}_1 : P_{X_1^{(1)}} \neq P_{X_1^{(2)}}$. To this end, Anderson, Hall and Titterington [2] proposed the test statistic

$$T_n = (n_1+n_2) \int_{\mathbb{R}^d} \left(\widehat{f}_1(x) - \widehat{f}_2(x)\right)^2 dx,$$

where $\widehat{f}_j = n_j^{-1}\sum_{k=1}^{n_j} K(x - X_k^{(j)}),\ j = 1,2,$ are kernel density estimators with fixed bandwidth. Alternatively, Alba Fernández, Jiménez-Gamero and Muñoz-García [1] suggested a characteristic function-based approach, i.e. they focused on a test statistic of the following form:

$$\widetilde{T}_n = (n_1+n_2) \int_{\mathbb{R}^d} |c_{n_1}(t) - c_{n_2}(t)|^2\, G(dt).$$

Here, G is a probability measure on $\mathbb{R}^d$ and $c_{n_j} = n_j^{-1}\sum_{k=1}^{n_j} e^{it'X_k^{(j)}},\ j = 1,2$. Additionally, they justified certain bootstrap methods to approximate the corresponding critical values. It would be useful to extend those results to the case of dependent observations, too.

A related topic for possible prospective investigations is the asymptotic theory for U-statistics of higher degree

$$U_n^{(k)} = \binom{n}{k} \sum_{\substack{j_1,\ldots,j_k=1\\ j_s\neq j_t \text{ for } s\neq t}}^{n} h(X_{j_1},\ldots,X_{j_k})$$

for weakly dependent data. If the kernel function h is first order degenerate, i.e. $\mathbb{E}h(x, X_2, \ldots, X_k) = 0, \ \forall\, x \in \mathbb{R}^d$, and the underlying random variables are i.i.d., then the limit is again a infinite weighted sum of χ^2-distributed random variables, cf. Lee [84], Section 3.2. One can conjecture that a similar result holds in the case of weakly dependent observations as well but this has to be verified in future work.

Bibliography

[1] Alba Fernández, V., Jiménez-Gamero, M. D. and Muñoz-García, J. (2008). A test for the two-sample problem based on empirical characteristic functions. *Comput. Statist. Data Anal.* **52**, 3730–3748.

[2] Anderson, N. H., Hall, P. and Titterington, D. M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *J. Multivariate Anal.* **50**, 41–54.

[3] Andrews, D. W. K. (1984). Non-strong mixing autoregressive processes. *J. Appl. Probab.* **21**, 930–934.

[4] Antille, A., Kersting, G. and Zucchini, W. (1982). Testing Symmetry. *J. Amer. Statist. Assoc.* **77**, 639–646.

[5] Arcones, M. A. and Giné, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.* **20**, 655–674.

[6] Babbel, B. (1989). Invariance principles for U-statistics and von Mises functionals. *J. Statist. Plann. Inference* **22**, 337–354.

[7] Bai, J. and Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *J. Bus. Econom. Statist.* **23**, 49–60.

[8] Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type. *Finance Stoch.* **2**, 41–68.

[9] Bartels, K. (1999). Tests zur Modellspezifikation in der nichtlinearen Regression. *Dissertation*.

[10] Berbee, H. C. P. (1979). *Random Walks with Stationary Increments and Renewal Theory.* Amsterdam: Mathematisch Centrum.

[11] Bickel, P. J. and Bühlmann, P. (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* **5**, 413–446.

[12] Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9**, 1196–1217.

[13] Bierens, H. J. (1982). Consistent model specification tests. *J. Econometrics* **20**, 105–134.

[14] BIERENS, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica* **58**, 1443–1458.

[15] BIERENS, H. J. AND PLOBERGER, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* **65**, 1129–1151.

[16] BILLINGSLEY P. (1968). *Convergence of Probability Measures*. New York-London-Sydney: John Wiley & Sons.

[17] BLOCK, H. W., LANGBERG, N. A. AND STOFFER, D. S. (1990). Time series models for non-Gaussian processes. In: *Topics in Statistical Dependence* (Block, H. W., Sampson, A. R. and Savits, T. H. eds.), Hayward, CA: Institute of Mathematical Statistics, 69–83.

[18] BORISOV, I. S. AND VOLODKO, N. V. (2008). Orthogonal series and limit theorems for canonical U- and V-statistics of stationary connected observations. *Siberian Adv. Math.* **18**, 242–257.

[19] BRADLEY, R. C. (1983). Approximation theorems for strongly mixing random variables. *Michigan Math. J.* **30**, 69–81.

[20] BROCKWELL, P. J. AND DAVIS, R. A. (1991). *Time Series: Theory and Methods*. New York: Springer.

[21] BÜHLMANN, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3**, 123–148.

[22] CARLSTEIN, E. (1988). Degenerate U-statistics based on non-independent observations. *Calcutta Statist. Assoc. Bull.* **37**, 55–65.

[23] CHEN, H. AND ROMANO, J. P. (1999). Bootstrap-assisted goodness-of-fit tests in the frequency domain. *J. Time Ser. Anal.* **20**, 619–654.

[24] CHEN, Y.-T., CHOU, R. Y. AND KUAN, C.-M. (2000). Testing time reversibility without moment restrictions. *J. Econometrics* **95**, 199–218.

[25] CHEN, X. AND WHITE, H. (1998). Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications. *Econometric Theory* **14**, 260-284.

[26] COULON-PRIEUR, C. AND DOUKHAN, P. (2000). A triangular central limit theorem under a new weak dependence condition. *Statist. Probab. Lett.* **47**, 61–68.

[27] DARLING, D. A. (1955). The Cramér-Smirnov test in the parametric case. *Ann. Math. Statist.* **26**, 1-20.

[28] DAUBECHIES, I. (1988). Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **41**, 909–996.

[29] Dedecker, J. and Doukhan, P. (2003). A new covariance inequality and applications. *Stochastic Process. Appl.* **106**, 63–80.

[30] Dedecker, J., Doukhan, P., Lang, G., León R., J. R., Louhichi, S. and Prieur, C. (2007). *Weak Dependence: with Examples and Applications.* New York: Springer.

[31] Dedecker, J. and Merlevède, F. (2006). Inequalities for partial sums of Hilbert-valued dependent sequences and applications. *Math. Methods Statist.* **15**, 176-206.

[32] Dedecker, J. and Merlevède, F. (2007). The empirical distribution function for dependent variables: asymptotic and nonasymptotic results in $\mathbb{L}^p$. *ESAIM Probab. Stat.* **11**, 102-114.

[33] Dedecker, J. and Prieur, C. (2004). Coupling for τ-dependent sequences and applications. *J. Theoret. Probab.* **17**, 861–885.

[34] Dedecker, J. and Prieur, C. (2004a). Couplage pour la distance minimale. *C. R. Acad. Sci. Paris, Ser. I* **338**, 805–808.

[35] Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields* **132**, 203–236.

[36] Dedecker, J. and Prieur, C. (2007). An empirical central limit theorem for dependent sequences. *Stochastic Process. Appl.* **117**, 121-142.

[37] Dedecker, J. and Rio, E. (2000). On the functional central limit theorem for stationary processes. *Ann. Inst. H. Poincaré Probab. Statist.* **36**, 1–34.

[38] Dehling, H. (1986). Almost sure approximations for U-statistics. In: *Dependence in Probability and Statistics* (Eberlein, E., Taqqu, M. S. eds.), Boston, MA: Birkhäuser, 119–135.

[39] Dehling, H., Denker, M. and Philipp, W. (1986). A bounded law of the iterated logarithm for Hilbert space valued martingales and its application to U-statistics. *Probab. Theory Related Fields* **72**, 111–131.

[40] Dehling, H. and Mikosch, T. (1994). Random quadratic forms and the bootstrap for U-Statistics. *J. Multivariate Anal.* **51**, 392–413.

[41] Dehling, H. and Taqqu, M. S. (1991). Bivariate symmetric statistics of long-range dependent observations. *J. Statist. Plann. Inference* **28**, 153–165.

[42] Dehling, H. and Wendler, M. (2010). Central limit theorem and the bootstrap for U-statistics of strongly mixing data. *J. Multivariate Anal.* **101**, 126–137.

[43] Denker, M. (1982). Statistical decision procedures and ergodic theory. In: *Ergodic Theory and Related Topics* (Michel, H. ed.), Berlin: Akademie-Verlag, 35–47.

[44] Denker, M. (1985). *Asymptotic Distribution Theory in Nonparametric Statistics.* Braunschweig: Vieweg.

[45] Dewan, I. and Prakasa Rao, B. L. S. (2001). Asymptotic normality for U-statistics of associated random variables. *J. Statist. Plann. Inference* **97**, 201-225.

[46] de Jong, R. M. (1996). The Bierens test under data dependence. *J. Econometrics* **72**, 1–32.

[47] de Wet, T. (1987). Degenerate U- and V-statistics. *South African Statist. J.* **21**, 99-129.

[48] de Wet, T. and Randles, R. H. (1987). On the effect of substituting parameter estimators in limiting χ^2 U and V statistics. *Ann. Statist.* **15**, 398–412.

[49] de Wet, T. and Venter, J. H. (1973). Asymptotic distributions for quadratic forms with applications to tests of fit. *Ann. Statist.* **1**, 380–387.

[50] Doukhan, P. (1994). *Mixing. Properties and Examples.* New York: Springer.

[51] Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84**, 313-342.

[52] Doukhan, P. and Neumann, M. H. (2008). The notion of ψ-weak dependence and its applications to bootstrapping time series. *Probab. Surv.* **5**, 146-168.

[53] Doukhan, P. and Wintenberger, O. (2008). Weakly dependent chains with infinite memory. *Stochastic Process. Appl.* **118**, 1997-2013.

[54] Duflo, M. (1997). *Random Iterative Models.* Berlin: Springer.

[55] Dunford, N. and Schwartz, J. T. (1988). *Linear Operators, Part II, Spectral Theory. Self Adjoint Operators in Hilbert Space.* New York: John Wiley & Sons.

[56] Eagleson, G. K. (1979). Orthogonal expansions and U-statistics. *Austral. J. Statist.* **21**, 221–237.

[57] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1–26.

[58] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.

[59] Escanciano, J. C. (2006). Goodness-of-fit tests for linear and nonlinear time series models. *J. Amer. Statist. Assoc.* **101**, 531–541.

[60] Escanciano, J. C. (2007). Model checks using residual marked empirical processes. *Statist. Sinica* **17**, 115–138.

[61] Escanciano, J. C. and Jacho Chávez, D. T. (2010). Approximating the critical values of Cramér-von Mises tests in general parametric conditional specifications. *Comput. Statist. Data Anal.* **54**, 625-636.

[62] Fan, Y. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. *Econometric Theory* **14**, 604–621.

[63] Fan, Y. and Li, Q. (1997). A consistent nonparametric test for linearity of AR(p) models. *Econom. Lett.* **55**, 53–59.

[64] Fan, Y. and Li, Q. (1999). Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *J. Nonparametr. Statist.* **10**, 245–271.

[65] Fan, Y. and Li, Q. (2000). Consistent model specification tests: kernel-based tests versus Bierens' ICM tests. *Econometric Theory* **16**, 1016–1041.

[66] Fan, Y. and Ullah, A. (1999). On goodness-of-fit tests for weakly dependent processes using kernel method. *J. Nonparametr. Statist.* **11**, 337–360.

[67] Feuerverger, A. and Mureika, R. A. (1977). The empirical characteristic function and its applications. *Ann. Statist.* **5**, 88–97.

[68] Ghosh, B. K. and Huang, W.-M. (1991). The power and optimal kernel of the Bickel-Rosenblatt test for goodness of fit. *Ann. Statist.* **19**, 999–1009.

[69] Gregory, G. G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.* **5**, 110-123.

[70] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications.* New York: Springer.

[71] Henze, N, Klar, B. and Meintanis, S. G. (2003). Invariant tests for symmetry about an unspecified point based on the empirical characteristic function. *J. Multivariate Anal.* **87**, 275–297.

[72] Hjellvik, V., Yao, Q. and Tjøstheim, D. (1998). Linearity testing using local polynomial approximation. *J. Statist. Plann. Inference* **68**, 295–321.

[73] Hodgson, D. J., Linton, O. and Vorkink, K. (2002). Testing the capital asset pricing model efficiently under elliptical symmetry: a semiparametric approach. *J. Appl. Econometrics* **17**, 617–639.

[74] Huang, W. and Zhang, L.-X. (2006). Asymptotic normality for U-statistics of negatively associated random variables. *Statist. Probab. Lett.* **76**, 1125–1131.

[75] Ignaccolo, R. (2004). Goodness-of-fit tests for dependent data. *Nonparametr. Statist.* **16**, 19–38.

[76] Jensen, M. B. and Lunde, A. (2001). The NIG-S&ARCH model: a fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. *Econom. J.* **4**, 319–342.

[77] Jiménez-Gamero, M. D. Muñoz-García, J. and Pino-Mejías, R. (2003). Bootstrapping parameter estimated degenerate U and V statistics. *Statist. Probab. Lett.* **61**, 61–70.

[78] Jose, K. K., Tomy, L. and Sreekumar, J. (2008). Autoregressive processes with normal-Laplace marginals. *Statist. and Probab. Lett.* **78**, 2456–2462.

[79] Knopp, K. (1956). *Infinite Sequences and Series.* New York: Dover Publications.

[80] Koroljuk, V. S. and Borovskich, Y. V. (1994). *Theory of U-Statistics.* Dordrecht: Kluwer Academic Publishers Group.

[81] Koul, H. L. and Stute, W. (1999). Nonparametric model checks for time series. *Ann. Statist.* **27**, 204–236.

[82] Kreiss, J.-P., Neumann, M. H. and Yao, Q. (2008). Bootstrap tests for simple structures in nonparametric time series regression. *Stat. Interface* **1**, 367–380.

[83] Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Ann. Statist.* **27**, 386–404.

[84] Lee, A. J. (1990). *U-statistics: Theory and Practice.* New York: Dekker.

[85] Lee, W.-M. (2007). Robust M tests using kernel-based estimators with bandwidth equal to sample size. *Econom. Lett.* **96**, 295–300.

[86] Leucht, A. and Neumann, M. H. (2009). Consistency of general bootstrap methods for degenerate U- and V-type statistics, *J. Multivariate Anal.* **100**, 1622–1633.

[87] Li, X. and Morris, J. M. (1991). On measuring asymmetry and the reliability of the skewness measure. *Statist. Probab. Lett.* **12**, 267–271.

[88] Li, Q. and Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *J. Econometrics* **87**, 145–165.

[89] Milhøj, A. (1985). The Moment Structure of ARCH processes. *Scand. J. Statist.* **12**, 281–292.

[90] Moore, D. S. and Spruill, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Ann. Statist.* **3**, 599–616.

[91] Munk, A., Stockis, J.-P., Valeinis J. and Giese, G. (2010). Neyman smooth goodness-of-fit tests for the marginal distribution of dependent data. *Ann. Inst. Stat. Math.*, DOI 10.1007/s10463-009-0260-2.

[92] Neumann, M. H. and Paparoditis, E. (2000). On bootstrapping L_2-type statistics in density testing. *Statist. Probab. Lett.* **50**, 137–147.

[93] Neumann, M. H. and Paparoditis, E. (2005). Goodness-of-fit tests for Markovian time series models. Technical Report 16/2005, Dept. Mathematics and Statistics, Univ. Cyprus.

[94] Neumann, M. H. and Paparoditis, E. (2008). Goodness-of-fit tests for Markovian time series models: central limit theory and bootstrap approximations. *Bernoulli* **14**, 14–46.

[95] Ord, J. K. (1968). The discrete Student's t distribution. *Ann. Math. Statist.* **39**, 1513–1516.

[96] Pemberton, J. and Tong, H. (1981). A note on the distributions of nonlinear autoregressive stochastic models. *J. Time Ser. Anal.* **2**, 49–52.

[97] Rachev, S. T. (2003). *Handbook of Heavy Tailed Distributions in Finance.* Amsterdam: Elsevier.

[98] Ramsey, J. B. and Rothman, P. (1996). Time irreversibility and business cycle asymmetry. *J. Money, Credit, Banking* **28**, 1–21.

[99] Rosenblatt, M. (1980). Linear processes and bispectra. *J. Appl. Probab.* **17**, 265–270.

[100] Shao, J. and Tu, D. S. (1995). *The Jackknife and Bootstrap.* New York: Springer.

[101] Shao, Q. M. and Yu, H. (1993). Bootstrapping the sample means for stationary mixing sequences. *Stochastic Process. Appl.* **48**, 175–190.

[102] Shao, X. and Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35**, 1773–1801.

[103] Staudte, R. G. and Sheather, S. J. (1990). *Robust Estimation and Testing.* New York: John Wiley & Sons.

[104] Stinchcombe, M. B. and White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* **14**, 295–325.

[105] Stute, W., Gonzáles Manteiga, W. and Presedo Quindimil, M. (1993). Bootstrap based goodness-of-fit tests. *Metrika* **40**, 243–256.

[106] Taufer, E. and Leonenko, N. (2009). Simulation of Lévy-driven Ornstein-Uhlenbeck processes with given marginal distribution. *Comput. Statist. Data Anal.* **53**, 2427–2437.

[107] USHAKOV, N. G. (1999). *Selected Topics in Characteristic Functions*. Utrecht: VSP.

[108] WEISS, G. (1975). Time-reversibility of linear stochastic processes. *J. Appl. Probab.* **12**, 831–836.

[109] WOJTASZCZYK, P. (1997). *A Mathematical Introduction to Wavelets.* Cambridge: Cambridge University Press.

[110] WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102**, 14150–14154.

[111] WU, W. B. AND MIN, W. (2005). On linear processes with dependent innovations. *Stochastic Process. Appl.* **115**, 939–958.

Ehrenwörtliche Erklärung

Hiermit erkläre ich,

- dass mir die Promotionsordnung der Fakultät bekannt ist,
- dass ich die Dissertation selbst angefertigt habe, keine Textabschnitte oder Ergebnisse eines Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel, persönliche Mitteilungen und Quellen in meiner Arbeit angegeben habe,
- dass ich die Hilfe eines Promotionsberaters nicht in Anspruch genommen habe und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,
- dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder eine andere wissenschaftliche Prüfung eingereicht habe,
- dass ich bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskriptes nicht von Dritten unterstützt wurde,
- dass ich weder die gleiche noch eine in wesentlichen Teilen ähnliche bzw. eine andere Abhandlung bereits an einer anderen Hochschule als Dissertation eingereicht habe.

Jena, 2010-12-14 Anne Leucht